# A Structural Investigation of Bacterial Twin-Arginine Translocation (Tat) Complexes by Single-Particle Electron Microscopy

Daniel Kaevon Beck BSc

A thesis submitted for the degree of Doctor of Philosophy

University of Warwick

Department of Life Sciences

# Table of Contents

# List of Figures

# List of Tables



# List of Equations

# Acknowledgements

I would first like to thank my supervisors Professor Colin Robinson and Dr Corinne J Smith for offering me the chance to undertake this research, and for their continual help and guidance throughout this endeavour.

I would also like to thank Ian Hands-Portman for his training, assistance and advice in the EM suite throughout my research. His help and input has been invaluable.

I would like to thank Professor Jan Maarten van Dijl and Carmine Monteferrante for the provision of isolated *B. subtilis* membranes.

Many thanks to all the members of the Robinson lab: to Jacopo, Anna, Chao and Roshani for being such good friends and providing such an enjoyable and supportive atmosphere to work in, and especially to Nishi Vasisht for her tireless efforts and generally looking after all of us over these years! Also thanks to Dr James Barnett for always finding time to listen to my ramblings and providing much sound advice.

I thank my beautiful fiancée Karina for her constant love and support throughout this undertaking. Without her I never would have emerged from the lab.

I thank my whole family for their love, understanding and contributions both financial and emotional.

# Declaration

The work presented in this thesis is original, and was conducted by the author, unless otherwise stated, under the supervision of Professor Colin Robinson and Dr Corinne J Smith.

None of the work presented in this thesis has been submitted previously for another degree.

This research was funded by the BBSRC.

The EM work in the Imaging Suite and the University of Warwick was supported by the Wellcome Trust (grant ref: 055663/Z/98/Z).

All sources of information have been acknowledged by means of reference.

Collaborative work has been carried out contributing towards material presented in this thesis. The extent of the collaborative work is indicated in the relevant chapters (Chapter 4). In summary, collaborative work was carried out with the following:

Jacopo Baglieri, Department of Life Sciences, University of Warwick

Isolated bacterial membranes were provided by others as indicated in the text where relevant (see Chapter 5).

Part of this work has been published as follows:

**Baglieri, J., Beck, D., Vasisht, N., Smith, C., and Robinson, C.** (2011). Structure of the TatA paralog, TatE, suggests a structurally homogeneous form of Tat protein translocase that transports folded proteins of differing diameter. *J. Biol. Chem.*

# Summary

The Twin arginine translocase (Tat) pathway was first characterised in chloroplast thylakoid membranes in the late 1990s. It has since been identified in the plasma membranes of both Gram-positive and Gram-negative bacteria. Substrates of this transport system contain a critical twin-arginine motif within their cleavable N-terminal signal sequence and the majority are large co-factor containing proteins.

There is now considerable evidence that Tat systems can transport such globular proteins in a fully folded state. The minimal components required for transport in E.coli are TatA, TatB and TatC; these three integral membrane proteins are thought to form an active translocon. In Bacillus subtilis only TatA and TatC subunits are present, with TatA acting in a bifunctional manner to replace TatB. Little structural information is known about these multimeric integral membrane protein complexes due to the inherent difficulty in purifying them and their compositional variability.

Complexes formed by *B. subtilis* TatAd and TatAyCy and *E. coli* TatE were investigated by single-particle EM analysis. An image processing protocol was developed to analyse and separate out individual Tat complexes based on their size. Using this method 3D electron density maps were generated of TatAd and TatE, which appear as small, ring-shaped complexes. Unlike *E. coli* TatA complexes, that have been shown to vary widely in size, those observed here appear small and homogeneous. These data conflict with the widely accepted 'size-fitting pore' model of Tat mediated translocation and rather support the alternative transient coalescent model.

Additionally the first structural characterisation of a TatA-type mutant protein was performed revealing a dramatic polymerisation phenotype and indicating a primary role for the N-terminus in forming protein-protein interactions.

## Abbreviations

| | |
|---|---|
| Å | Ångströms |
| APH | Amphipathic Helix |
| APS | Ammonium persulphate |
| ATP | Adenosine triphosphate |
| BN-PAGE | Blue Native PAGE |
| BSA | Bovine serum albumin |
| C- | Carboxy terminus |
| C | Concentrate |
| CC | Cross-correlation |
| CCC | Cross-correlation Coefficient |
| CCD | Charge-coupled Device |
| CD | Circular Dichroism |
| CMC | Critical Micelle Concentration |
| CTF | Contrast Transfer Function |
| *Δ* | Delta (gene deletion). |
| DDM | *n*-Dodecyl-*ß*-D-maltoside |
| DHFR | Dihydrofolate reductase |
| DMSO | Dimethyl sulphoxide |
| E | Elution fraction |
| ECL | Enhanced chemifluoresence |
| EDTA | Ethylenediaminetetraacetic acid |
| EM | Electron microscopy |
| F | Filtrate |
| FFT | Fast Fourier Transform |
| FSC | Fourier Shell Correlation |

| | |
|---|---|
| FT | Flow-through fraction |
| GF | Gel filtration |
| GUI | Graphical User Interface |
| *his / h* | Hexahistidine tag |
| HAC | Hierarchical Ascendance Classification |
| HRP | Horseradish peroxidise |
| I | Importance Image |
| IMVs | Inverted inner membrane vesicles |
| kDa | Kilodaltons |
| LB | Luria Bertani medium |
| M | Membrane fraction |
| mAU | Milli-absorbance unit |
| MDa | Megadaltons |
| MDS | Minimal Dose Settings |
| MSA | Multivariate Statistical Analysis |
| MWCO | Molecular Weight Cut Off |
| N- | Amino terminus |
| nm | Nanometres |
| NMR | Nuclear Magnetic Resonance |
| NTA | Nitrilotriacetic acid |
| OD | Optical density |
| P | Periplasmic fraction |
| PAGE | Polyacrylamide gel electrophoresis |
| φ | Phi (Euler angle) |
| PBS | Phosphate buffered saline |
| PDC | Protein Detergent Complexes |
| PMF | Proton motive force |

| | |
|---|---|
| ψ | Psi (Euler angle) |
| PVDF | Polyvinylidene fluoride membrane |
| R | Reconstituted Image |
| RCT | Random Conical Tilt |
| RNA | Ribonucleic acid |
| rpm | Revolutions per minute |
| σ | Sigma (Standard Deviation above the mean density) |
| S/N | Signal to Noise |
| SIRT | Simultaneous Iterative Reconstruction Technique |
| SDS | Sodium dodecyl sulphate |
| SDVAR | Standard Deviation of Variance |
| Sec | General secretory pathway |
| SPIDER | System for Processing Image Data from Electron microscopy and Related fields |
| SPIRE | SPIDER Reconstruction Engine |
| SRP | Signal recognition particle |
| *Strep / s* | Strep II tag |
| T | Total pooled protein |
| Tat | Twin arginine translocation |
| TCA | Trichloroacetic acid |
| TEMED | Tetramethylethylenediamine |
| θ | Theta (Euler angle), Tilt. |
| TMAO | Trimethylamine-*N*-oxide |
| TMH | Transmembrane Helix |
| TorA | Trimethylamine-*N*-oxide reductase |
| $V_o$ | Void Volume |
| v/v | Volume per volume |

| | |
|---|---|
| W | Wash fraction |
| WT | Wild type |
| w/v | Weight per volume |

## SPIDER operations

Presented here is a comprehensive list of all SPIDER operations used in this investigation for future reference purposes. See Appendix for their usage within the listed batch files.

See www.wadsworth.org/spider_doc/spider/docs/operations_doc.html for the full manual of SPIDER operations, where an explanation of the function of each can be found.

**AP MS**

**AP SA**

**AP SH**

**AP SR**

**AR**

**AS DC**

**AS R**

**BP 32F**

**BP CG**

**BP CG3**

**BP RP**

**CA S**

**CA SM**

**CA SRA**

**CA SRD**

**CC C**

**CG**

**CL HC**

**CL HD**

**CL HE**

**CP**

**CP FROM RAW**

**DE**

**DE A**

**DO**

**DOC COMBINE**

**DOC RAN**

**DOC REN**

**DOC SORT**

**DOC SUBTRACT**

**EN**

**EN D**

**FQ**

**FQ NP**

**FR G**

**FR L**

**FS**

**FT**

**HD D**

**IF**

**IF GOTO**

**IF THEN**

**LB#**

**MA**

**MD**

**MM C**

**MN S**

**MO**

**MU**

**PJ 3Q**

**RF 3**

**RO I**

**RT SQ**

**SD**

**SD C**

**SD E**

**SD IC NEW**

**SH F**

**TF CT**

**TF ED**

**TH M**

**UD**

**UD E**

**UD FIND**

**UD ICE**

**UD N**

**VM**

**WI**

# *Chapter 1*

# *Introduction*

## 1.1. Protein translocation in bacteria

The compartmentalisation of cellular components is one of the most vital features within all living organisms. Without the presence of segregating membranes within even the simplest of cells, vital biochemical reactions would never occur; energy generation within cells relies on the presence of a semi-permeable membrane to produce an ion gradient and drive production of ATP. Various toxic and otherwise damaging cellular components are carefully sequestered within membrane enclosed vesicles so other biochemical processes can proceed unhindered. As such any movement of large biomolecules across these membranes must be conducted with specificity and efficiency so as not to disrupt ion gradients or compartmental compositions.

Although bacteria lack the more complex compartmentalisation present in eukaryotic cells, in order to interact with its environment the bacterial cell must allow biomolecules to enter the cytoplasm, and secrete its own proteins. In both cases this requires trafficking across the plasma membrane, without compromising its integrity.

Bacterial prokaryotes can be classified into two distinct groups based on their retention of the Gram-stain dye; those that retain the dye are termed Gram-positive and those that do not are termed Gram-negative. The difference in dye-retention, and hence the differentiating feature between these groups, is due to a variation in cell wall composition, as shown in Figure 1.1.1 The Gram-negative bacterium presents the more complex structure; with the cytoplasmic membrane surrounded by a thin layer of peptidoglycan, followed by an outer membrane enclosing the periplasmic space. In contrast the Gram-positive bacterium has only a single membrane surrounded by a much thicker cell wall of peptidoglycan and a co-anionic polymer.

The bacterial secretome is vast, with ~20 % of all bacterial proteins synthesised within the cytoplasm being translocated across the cytoplasmic membrane; for example, *Escherichia coli* translocates over 250 proteins, and it is believed that *Bacillus subtilis* has a potential secretome of 300 proteins (Tjalsma *et al.*, 2000). The majority of these secreted proteins use the well characterised Sec (general secretory) pathway via a threading mechanism that has been studied in much detail (Berg *et al.*, 2004; Jarvis and Robinson, 2004; Manting and Driessen, 2000). Other, more specialised protein translocation systems are also found in prokaryotes (Backert and

Meyer, 2006; Christie *et al.*, 2005; Cornelis, 2006; Henderson *et al.*, 2004; Johnson *et al.*, 2006), mostly in the outer-membrane of Gram-negative bacteria.

A small proportion of secreted proteins use the alternative Tat (twin arginine translocation) pathway. The Tat system has the unique and fascinating ability to translocate fully folded proteins across cell membranes. This study presents an investigation into the structure of bacterial Tat complexes, focusing on *E. coli* and *B. subtilis* as model Gram-negative and Gram-positive bacteria, respectively.

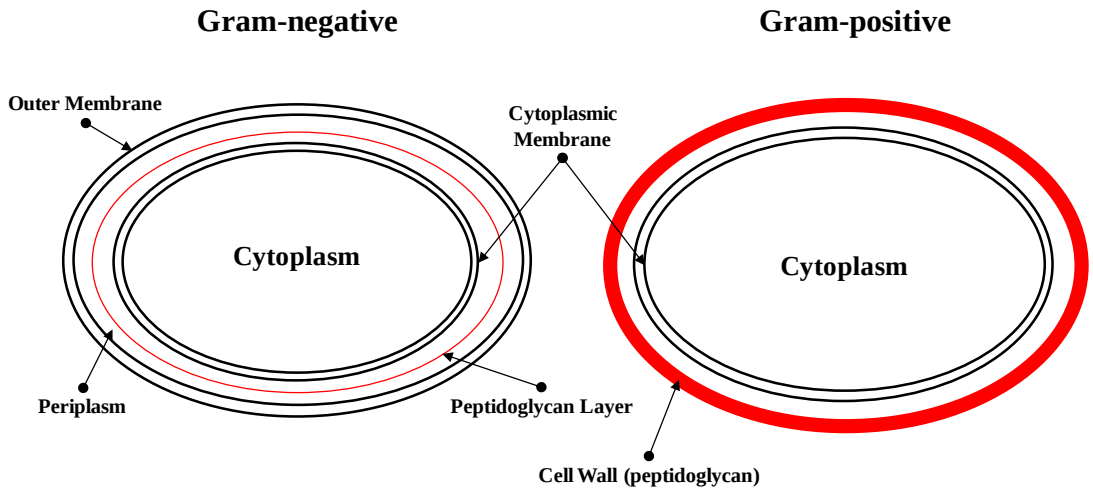



**Figure 1.1.1 The Cell walls of Gram-negative and Gram-positive bacteria**
A schematic representation of the bacterial cell wall for both Gram-stain types is shown. Gram-positive bacteria have a very thick layer of peptidoglycan and co-anionic polymer.

## 1.2. The General Secretory (Sec) pathway

The Sec pathway is responsible for the translocation of most bacterial proteins. It is found in all domains of life, both Gram-positive and Gram-negative bacteria (Mori, Hiroyuki and Ito, 2001), arachaea, the thylakoid membrane of plants and in the eukaryotic endoplasmic reticulum (Muñiz *et al.*, 2001). In all organisms the Sec system works in a similar way, translocating proteins in an unfolded state via a threading mechanism utilising an ATP-dependent motor protein and a membrane localised pore.

This pathway has been very well characterised and extensively reviewed (Jarvis and Robinson, 2004; Manting and Driessen, 2000; Natale *et al.*, 2008; Pugsley, 1993).

The Sec protein conducting channel (PCC) consists of three integral membrane proteins (SecY, SecE and SecG) which form a membrane spanning heterotrimeric SecYEG pore; the x-ray crystal structure of which reveals a dimer with an hourglass shape, ranging from ~ 25 Å to ~ 4 Å wide (Berg *et al.*, 2004). This pore is plugged by both a periplasmic loop of SecY (Li *et al.*, 2007), and a gasket-like seal formed by the pore ring itself, that prevents the flow of small molecules during translocation (Park and Rapoport, 2011).

Secretory proteins can be guided to the Sec PCC by the specific chaperone protein SecB, that prevents them from mis-folding in the cytoplasm (Driessen, 2001); in other cases the substrate is bound by more general chaperones such as GroEL. SecB binds to the substrate after it leaves the ribosome (Kumamoto and Francetic, 1993) and targets to the Sec PCC via an interaction with SecA (Hartl *et al.*, 1990). SecA is a hydrophilic protein of 102 kDa that is associated peripherally with the Sec PCC where it acts as an ATP-driven motor, providing the force required for translocation.

This system is also involved in the insertion of membrane proteins into the bacterial cytoplasmic membrane (Luirink *et al.*, 2005). In bacteria this involves targeting to the Sec system in a co-translational manner, via the secondary signal recognition particle (SRP) pathway (Luirink and Sinning, 2004; Valent, 2001). The SRP is a ribonucleoprotein that binds to the signal-peptide as it emerges from the ribosome, forming the ribosome nascent chain (RNC)-SRP complex (Nagai *et al.*, 2003), that is then targeted in its entirety to the Sec PCC. Some bacterial secretory proteins use this

route, however, in the eukaryotic ER it is utilised by virtually all secretory proteins. Recent cryo-EM studies have been able to show in sub-nanometre detail the structure of the Sec PCC with a bound ribosome, stalled in the act of translation (Frauenfeld *et al.*, 2011).

## 1.3. The Twin-Arginine Translocation (Tat) pathway

### 1.3.1. Introduction

The alternative Tat pathway was first characterised in plant chloroplasts (where it is now termed cpTat) in the early 1990s, this system was found to require a proton motive force for activity (Klösgen *et al.*, 1992; Mould and Robinson, 1991) but was not reliant on ATP or other stromal factors (Cline, K. *et al.*, 1992; Mould *et al.*, 1991). The cpTat pathway was later shown to be capable of translocating proteins in a folded state (Clark and Theg, 1997; Creighton *et al.*, 1995; Hynds *et al.*, 1998), arguably the most dramatic and unique feature ubiquitous to Tat systems. The main differences between the Tat and Sec translocation pathways are highlighted in Figure 1.3.1

This unique system has been the focus of much research attention, the scope of which has been reviewed periodically (Fröbel *et al.*, 2012; Lee *et al.*, 2006; Müller, Matthias and Klösgen, 2005; Robinson and Bolhuis, 2004; Robinson *et al.*, 2011). Some of the more recent studies have begun to focus on potential uses of Tat within the biopharmaceuticals sector (Branston *et al.*, 2012; Matos *et al.*, 2012). The ability to secrete medically relevant proteins in a fully-folded state during industry scale purification could go a long way towards circumventing the issue of inclusion body formation; this remains a major stumbling block during the scale-up of many such experiments.

Despite many years of work, the current understanding of the Tat system remains limited in terms of its structure and function. This is due to the inherent difficulties in working with large, multimeric, membrane integral complexes and is most restricted by a lack of high-resolution structural data.

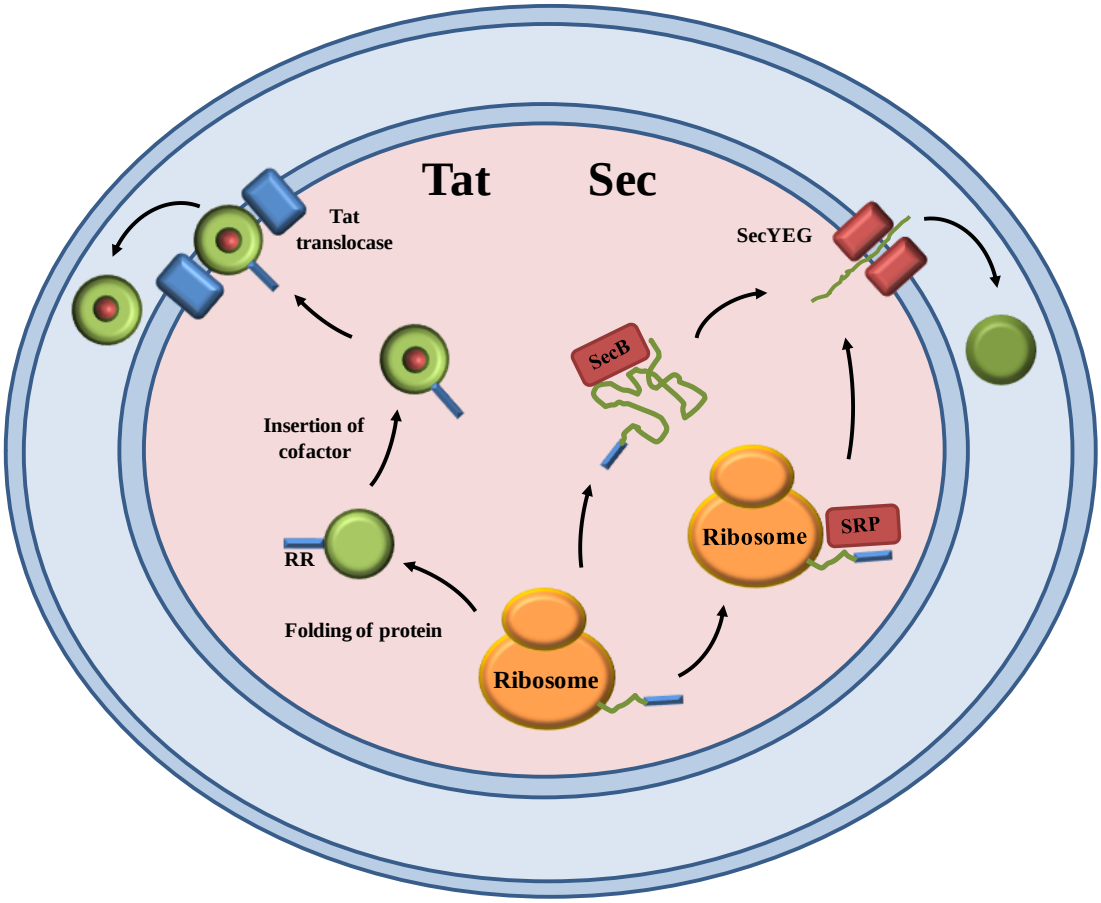



**Figure 1.3.1. Comparison of the Tat and Sec translocation pathways**
A schematic representation comparing the main features of the Sec and Tat pathways within *E. coli* is shown. The Tat pathway is shown on the left and the co-translational and post-translational (SRP) routes of the Sec pathway are shown on the right. Proteins are shown during export into the periplasm.

### 1.3.2. Tat genes

Within plants and Gram-negative bacteria such as *E. coli*, the Tat pathway requires three proteins for translocase activity. These were initially termed Tha4 (Walker *et al.*, 1999), Hcf106 (Settles *et al.*, 1997) and cpTatC (Cline and Mori, 2001) in plants, but are more recently termed cpTatA, cpTatB and cpTatC in reference to their bacterial counterparts. Accordingly, the bacterial homologues are named TatA, TatB and TatC (Bogsch, Erik G. *et al.*, 1998; Sargent *et al.*, 1998; Sargent *et al.*, 1999). All Gram-positive bacteria, with the exception of the actinomycetes (Schaerlaekens *et al.*, 2001), possess a minimal or AC-type Tat system where only TatA and TatC subunits are present (Barnett *et al.*, 2008; Jongbloed *et al.*, 2004; Pop *et al.*, 2002; Yen *et al.*, 2002).

The genes encoding TatABC in *E. coli* are found together in an operon along with a fourth *tat* gene, *tatD*; this encodes a soluble nuclease that is not required for translocation activity (Wexler *et al.*, 2000). Although the *tat* operon is constitutively expressed in *E. coli*, its expression is not essential for growth, except during anaerobic conditions when grown on minimal media (Santini *et al.*, 1998; Sargent *et al.*, 1998). *E. coli* also contains a TatA paralogue termed TatE, this is encoded separately by the monocistronic *tatE* gene and is unique to enterobacteria (Yen *et al.*, 2002), with the exceptions of the Gram-positive bacteria *Corynebacterium glutamicum* and *Corynebacterium efficiens* (Ikeda and Nakagawa, 2003).

In the Gram-positive model organism *Bacillus subtilis*, two minimal (AC-type) Tat pathways have been identified (Jongbloed *et al.*, 2004), organised in two operons; the first consists of *tatAd* and *tatCd* along with *phoD* which encodes the only identified substrate of this pathway, the second contains *tatAy* and *tatCy* with the lone substrate of this system, YwbN expressed elsewhere. A third *tatA*-type gene, *tatAc*, is expressed on its own and has no direct role in either pathway. A schematic representation displaying the organisation of the *tat* genes of *E. coli* and *B. subtilis* is shown below in Figure 1.3.2.

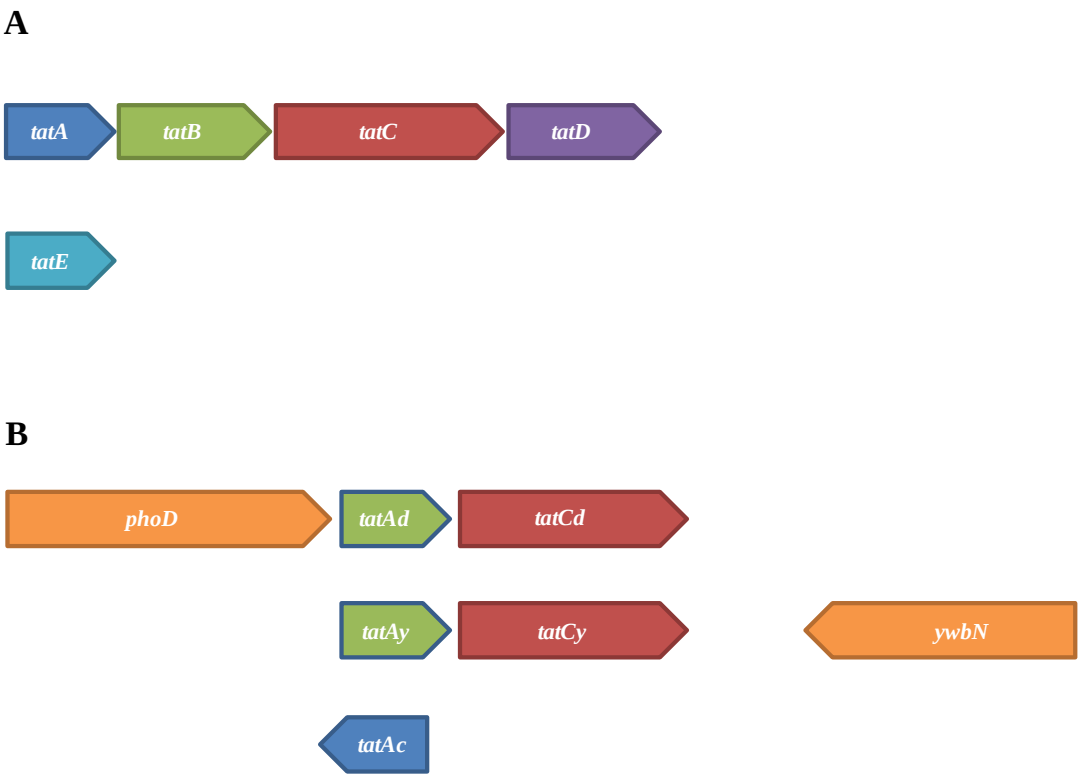



**Figure 1.3.2. The organisation of the *tat* genes from *E. coli* and *B. subtilis***
*A.* The *tatABC* genes are the minimal components required for translocation in *E. coli* and are found in an operon alongside the *tatD* gene. The *tatE* gene is a paralogue of *tatA* and is expressed elsewhere in the genome. *B.* Within the Gram-positive bacterium *B. subtilis* two minimal (AC-type) Tat pathways are present, organised in two operons. The first consists of the *tatAd* and *tatCd* genes along with the substrate *phoD*, which utilises this pathway. The *tatAy* and *tatCy* genes form the second operon, with the TatAyCy substrate YwbN being expressed elsewhere. The third *tatA*-type gene, *tatAc* is not expressed alongside any of the other *tat* genes and does not have a direct role in either pathway.

### 1.3.3. Tat subunits

The TatA subunit of *E. coli* consists of 89 amino acids (9.6 kDa); sequence analyses have predicted TatA-type proteins to consist of a single transmembrane (TM) spanning helix and a longer amphipathic helix (APH), joined together by a short hinge region (Porcelli *et al.*, 2002). These secondary structure predictions have been confirmed in TatAd using circular dichrosim (CD) and oriented CD (OCD) techniques (Lange *et al.*, 2007).

The TatA orthologues are thought to be arranged with the N-terminus exposed to the periplasm (Porcelli *et al.*, 2002), and the short, charged C-terminal tail trailing into the cytoplasm. However, this topology has been brought into question using thiol-reactive and accessibility experiments (Chan *et al.*, 2007; Gouffi *et al.*, 2004) leading to the proposal of a dual topology. Studies on *B. subtilis* TatAd using solid-state NMR have shown that the APH is oriented parallel to the lipid bilayer (Müller, Sonja D. *et al.*, 2007; Walther *et al.*, 2010), forming the L-shaped arrangement as shown in Figure 1.3.3. The TM domain was found to be tilted by 17° relative to planar bicelles, indicating that both the hinge region, and N-terminal portion of the APH are also lipid immersed. The APH was also noted to adopt a slanted orientation at the interface, modifying somewhat the membrane-parallel alignment indicated previously (Lange *et al.*, 2007). Similar experiments using solution-state methods on TatAd have indicated that the TMH consists of 13 residues and the APH is 23 residues long with a high level of flexibility in the C-terminal region (Hu *et al.*, 2010).

The TatE protein of *E. coli* is shorter than TatA at only 67 amino acids, but shares over 50% sequence identity with its paralogue (Sargent *et al.*, 1998), and as such is thought to have arose via a gene duplication (Yen *et al.*, 2002). Although TatE has shown functional overlap with TatA, in its ability to complement a Δ*tatA* mutant (Sargent *et al.*, 1999), it has largely been considered redundant as it is expressed at much lower levels than TatA (Jack, R. L. *et al.*, 2001). However, TatE does play a role in the production of biofilms (Heikkila *et al.*, 2001) and its absence has been noted to strongly retard the translocation of selected Tat substrates (Sargent *et al.*, 1998), suggesting a more specialised role within the Tat translocase.

The *E. coli* TatB subunit shares 25% sequence similarity with TatA, with much of this focused on the functionally significant N-terminus, predicted TM-span and APH (Hicks *et al.*, 2003). It is therefore predicted to adopt a similar membrane-integral topology to TatA (Bolhuis *et al.*, 2001), as shown below (Figure 1.3.3). However, TatB is significantly longer than TatA at 171 amino acids (18.4 kDa), as it possesses a long, flexible C-terminal tail that is not essential for its function (Lee *et al.*, 2002). The presence of TatB is essential for translocation of native *E. coli* substrates but artificial fusion-substrates have been produced that can somewhat tolerate its absence (Blaudeck *et al.*, 2005; Ize *et al.*, 2002). This suggests that residual TatB-like activity may be apparent in *E. coli* TatA, echoing the truly bifunctional TatA homologues of *B. subtilis* AC-type minimal Tat systems (Barnett *et al.*, 2008). Such bifunctionality can be induced in *E. coli* TatA by single amino acid substitutions to the extreme N-terminus (Barrett *et al.*, 2007; Blaudeck *et al.*, 2005), indicating a vital role for this region.

The largest subunits are the TatC-type proteins. *E. coli* TatC is 28.9 kDa and 258 amino acids long, it is predicted to consist of six TM spans oriented with the N and C termini facing the cytosol (Behrendt *et al.*, 2004). The predicted high helical content of TatCd has been confirmed using CD and OCD methods (Nolandt *et al.*, 2009); however, a tilted arrangement in the bilayer was also noted suggesting not all the helices may span the membrane. Although the presence of six TM helices is the most likely arrangement, a previous paper has suggested only four are present in TatC (Gouffi *et al.*, 2002). The interconnecting loop regions that protrude into the cytoplasm have been identified as important for protein function via several mutational analyses (Allen *et al.*, 2002; Barrett *et al.*, 2005; Buchanan *et al.*, 2002; Holzapfel *et al.*, 2007; Punginelli *et al.*, 2007).

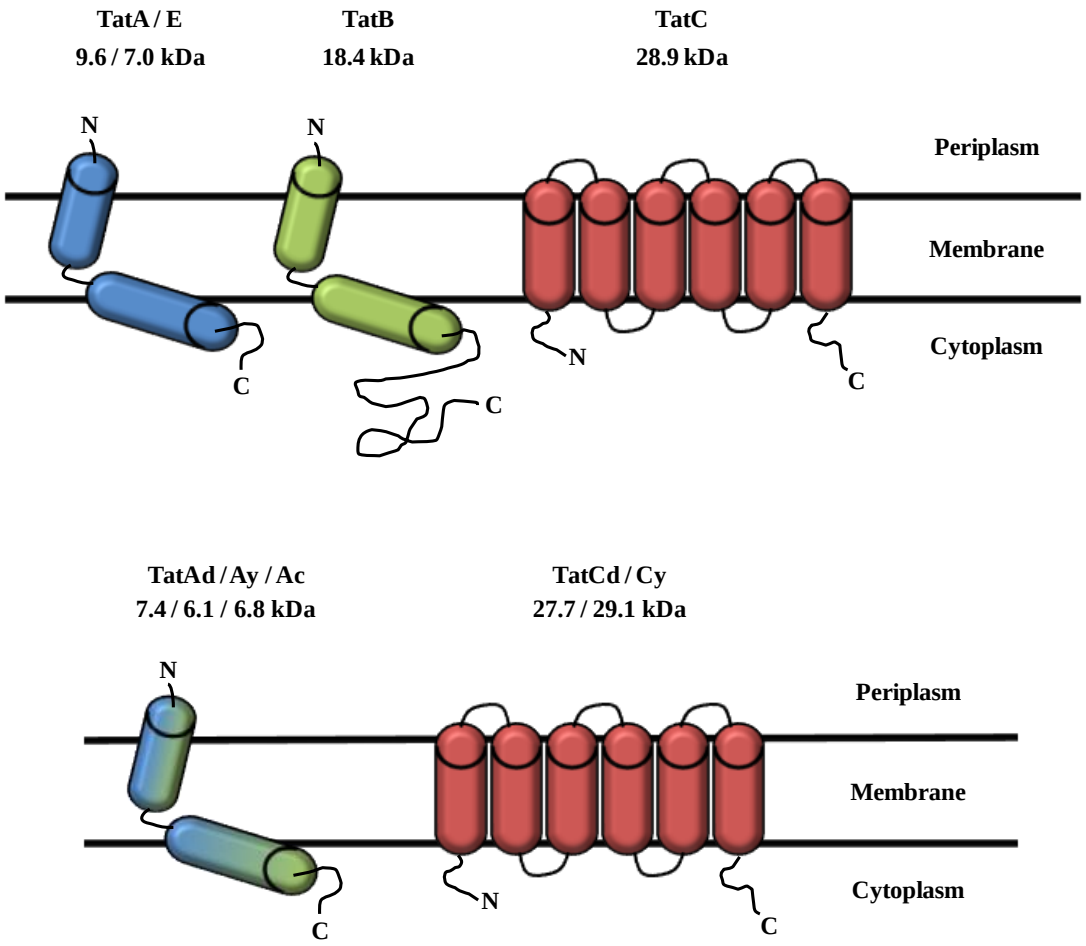



**Figure 1.3.3. Topology of the Tat proteins of *E. coli* and *B. subtilis***
A schematic representation of the *E. coli* (top) and *B. subtilis* (bottom) Tat subunits are shown. The TatA-type and TatB subunits consist of a single transmembrane (TM) helix joined to an amphipathic helix by a sort hinge region. The tilted arrangement within the membrane has been suggested by recent NMR studies into TatAd (Hu *et al.*, 2010; Walther *et al.*, 2010). TatB has a markedly longer C-terminal extension than the other single TM-span Tat subunits. The TatC-type subunits are made up of 6 TM-spans joined by short loops, with both termini arranged at the cytosolic face

### 1.3.4. Tat substrates

The folded state of Tat substrates is not just tolerated, but is a selective property for translocation via the Tat machinery. This requirement has been demonstrated through both *in vitro* and *in vivo* means using model substrates with attached Tat signal peptides; disulphide-bond formation was a prerequisite for translocation of alkaline phosphatase fusion proteins (DeLisa *et al.*, 2003; Panahandeh *et al.*, 2008), and cytochrome *c* could only be translocated after haem insertion (Sanders *et al.*, 2001). Earlier experiments in plant thylakoids utilising dihydrofolate reductase (DHFR) in the presence of folding-inductive methotrexate (Hynds *et al.*, 1998) have also demonstrated this ability to transport folded substrates. However, in plants the Tat system appears to be less selective and is capable of translocating unfolded proteins.

By way of an explanation for this folded-state requirement many Tat substrates are cofactor containing proteins, and as such undergo insertion of the cofactor prior to translocation. The correct folding of these apoproteins is reliant on cofactor insertion and subsequently is a prerequisite for translocation via Tat (Berks, B. C., 1996; Ilbert *et al.*, 2003; Sanders *et al.*, 2001; Santini *et al.*, 1998).

A range of cytosolic chaperones, also termed REMPs (Redox enzyme maturation proteins), have been identified for the *E. coli* Tat system. These effectively have the opposite role to those of the Sec system (SecB and SRP) as they prevent interaction with the translocase to allow folding and cofactor insertion to occur; this constitutes a form of proof-reading that acts independently of the Tat translocase. The best characterised proof-reading chaperones are TorD, DmsD, NapD, HyaE and HybE which bind to the signal peptides of TorA, DmsA, NapA, HyaA and HybO respectively (Dubini and Sargent, 2003; Genest *et al.*, 2006; Jack, Rachael L *et al.*, 2004; Oresnik *et al.*, 2001). This proof-reading ability has been well demonstrated for TorD and HybE, which have been shown to prevent premature targeting and translocation of Tat substrates (Jack, Rachael L *et al.*, 2004).

Interestingly some Tat substrates are seen to completely lack a signal peptide and are forced to use a hitch-hiking method (Rodrigue *et al.*, 1999) to target to the Tat translocase. These proteins exist as hetero-oligomers where only a single subunit

possesses a signal peptide. These include *E. coli* hydrogenases (Rodrigue *et al.*, 1999) and DmsB (Dimethyl sulfoxide subunit B) which must interact with its signal peptide containing heterodimer, DmsA (Sambasivarao *et al.*, 2000). These hitch-hiking proteins must be properly folded in order for the relevant protein-protein interactions to be made.

It has been suggested that monomeric Tat substrates that lack cofactors may exhibit rapid folding kinetics; hence they cannot be transported via Sec. Such substrates include AmiA and AmiC, which are required for normal cell wall biosynthesis. In *E. coli Δtat* mutants AmiA and AmiC are mis-localised resulting in the definitive *tat* phenotype with cells growing in filamentous chains (Bernhardt and de Boer, 2003; Ize, B. *et al.*, 2003) as normal cell division is retarded.

The majority of Tat substrates are soluble proteins but some remain anchored to the membrane following translocation, including the Rieske iron-sulphur protein (Aldridge *et al.*, 2008; Bachmann *et al.*, 2006; De Buck *et al.*, 2007; Hatzixanthis *et al.*, 2003). This suggests that some kind of lateral membrane insertion event can take place during translocation.

### 1.3.5. Tat signal peptides

The name of the Tat system is derived from the virtually invariant twin-arginine motif present in the N-terminal region of Tat signal peptides (Chaddock *et al.*, 1995). These signal peptides contain the highly conserved consensus motif S/T-R-R-x-F-L-K (where x represents any polar amino acid) (Berks, B. C., 1996; Cristóbal *et al.*, 1999), but are otherwise very similar in structure to the Sec signal peptide. In both cases they are divided into three regions; a basic n-region, a hydrophobic h-region, and a polar c-region (listed from N to C termini). The carboxy-terminal region contains a recognition sequence for cleavage of the signal peptide by signal peptidase I (Luke *et al.*, 2009; Yahr and Wickner, 2001). The organisation of the signal peptide for *E. coli* Tat substrate TorA is shown below in Figure 1.3.4.

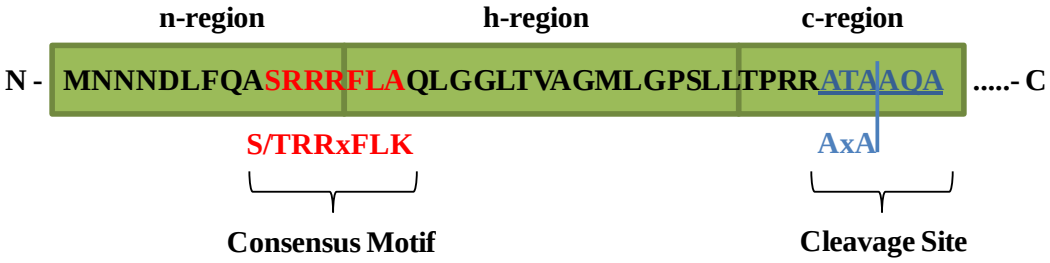



**Figure 1.3.5 Example of an *E. coli* Tat signal peptide**
The N-terminal signal peptide of TorA, an *E. coli* Tat substrate commonly used for translocation activity assays, is shown. Tat signal peptides consist of 3 domains; the positively charged n-region, the hydrophobic h-region and the polar c-region. The consensus motif is indicated along with the cleavage site for signal peptidase.

The importance of the twin-arginine motif is evident from the complete abolition of translocation induced by a dual substitution with lysine. Within plants the twin-arginine motif has been shown to be an absolute requirement (Chaddock *et al.*, 1995); however, substitution of a single arginine can be tolerated in bacteria (DeLisa *et al.*, 2002; Ize, Bérengère *et al.*, 2002; Mendel *et al.*, 2008; Stanley *et al.*, 2000), and single Arg containing sequences are found naturally (Hinsley *et al.*, 2001; Ignatova *et al.*, 2002; Molik *et al.*, 2001; Widdick *et al.*, 2008). Other residues within

the consensus motif also play a vital role in directing a substrate to the Tat pathway; however, the importance of the individual amino acids is often variable between substrates (Mendel *et al.*, 2008; Stanley *et al.*, 2000). It has also been noted that the same, or very similar, targeting signals apply to the Tat systems of both Gram-positive and Gram-negative bacteria.

More general properties inherent in these sequences have been identified as targeting constituents for the attached protein. The overall low hydrophobicity of the Tat signal peptides have been implemented in targeting to the Tat system; with hydrophobic mutations in the n-region (Ize, B. *et al.*, 2002) or h-region (Cristóbal *et al.*, 1999) resulting in Tat substrates being directed to the Sec system. The basic amino acid content of the c-region has also been highlighted as a Sec system avoidance motif (Blaudeck, N. *et al.*, 2003; Bogsch, E. *et al.*, 1997).

## 1.4. The structure of Tat complexes

Since the initial discovery of the Tat system structural investigations into these complexes have provided some of the most interesting and conflicting results within the field. The intrinsic difficulties associated with studying these large, multimeric, membrane-spanning complexes have all but ruled out an X-ray crystal structure, leaving electron microscopy, NMR and other spectroscopic techniques to fill in this large knowledge gap. Various Tat subunits have shown the ability to homo-oligomerise under both *in vitro* and *in vivo* conditions. However, determining which of these complexes represent biologically relevant units, and which are experimental artefacts, remains a very difficult task. Nevertheless, high resolution structural models of these complexes are needed to answer many of the questions within the field. Such data can both reinforce the conclusions of previous biochemical studies and provide much needed insights for new investigations into this unique system and its puzzling translocation mechanism.

### 1.4.1. Tat complexes in Gram-negative bacteria

Most early structural studies have focused on *E.coli*, where multimeric complexes of both TatABC and TatA have been reported (De Leeuw *et al.*, 2002; Oates *et al.*, 2005). The presence of TatA within the TatABC complex is somewhat contentious however, and TatBC complexes have been reported in plants (Cline and Mori, 2001) and bacteria (Behrendt *et al.*, 2007; Tarry *et al.*, 2009). As such, this heteromeric complex has been referred to as Tat(A)BC (Fröbel *et al.*, 2012) to standardise the nomenclature.

Although it has been shown that for full Tat-dependent protein export TatA, TatB and TatC subunits are required (Bogsch *et al.*, 1998; Sargent *et al.*, 1998; Weiner *et al.*, 1998) a number of complexes consisting of these subunits have been purified and characterised by BN-PAGE or gel-filtration. These multimeric complexes range from ~100-600 kDa in size, including an unknown contribution by detergent (Gohlke *et al.*, 2005; Oates *et al.*, 2005; Oates *et al.*, 2003; Sargent, F *et al.*, 2001). In most EM imaging experiments a cylindrical structure with a central cavity or channel has been observed leading to the commonly accepted "pore-forming unit" or translocon model

(Gohlke *et al.*, 2005; Oates *et al.*, 2003; Sargent, F *et al.*, 2001; Westermann *et al.*, 2006).

One of the earliest direct structural studies into the Tat system focused on complexes containing only TatA and TatB. Single-particle EM analysis of negatively stained micrographs revealed 2D average projections showing cylindrical double-ringed structures (~90 Å thick) (Sargent, F *et al.*, 2001). These appeared ~160 Å wide consisting of six or seven peripheral 35 Å globular densities and a ~65 Å wide central region. The complex was considerably larger than the ~50 Å distance required to span the membrane bilayer. However, the role or presence of TatB within these complexes has been called into question by conflicting results in later studies (Bolhuis *et al.*, 2001; Gohlke *et al.*, 2005; Oates *et al.*, 2005) leaving some to believe that the complex may represent an assembly intermediate or TatA homomeric complex.

It has been shown by co-immunoprecipitation experiments that TatC binds tightly with TatB within a purified TatABC complex and is required for the interaction of TatA with TatB (Bolhuis *et al.*, 2001). In this study the TatB and TatC subunits were found in a strict 1:1 ratio confirmed by the expression of a functional Tat(BC) chimera by a translational fusion. Although some TatA was tightly associated with TatB/C in variable amounts the vast majority did not co-purify, suggesting a weaker or transient interaction.

Tat(A)BC complexes vary in size between 360-700 kDa (Bolhuis *et al.*, 2001; Oates *et al.*, 2005; Oates *et al.*, 2003) depending on method of isolation. In the absence of TatA, *E. coli* TatB and TatC can associate to form large complexes (Behrendt *et al.*, 2007; McDevitt *et al.*, 2005; Orriss *et al.*, 2007), suggesting that these subunits form the core of the Tat(A)BC complex. However this TatBC complex is often found associated with small amounts of TatA (Bolhuis *et al.*, 2001; De Leeuw *et al.*, 2002; McDevitt *et al.*, 2005; Oates *et al.*, 2005), and the presence of TatA has been reported to stabilise the TatBC complex (Mangels *et al.*, 2005). It is broadly accepted that the Tat(A)BC complex acts as the signal-peptide recognition and binding site for Tat substrates.

More recently, work on purified TatBC complexes seen by BN-PAGE (Orriss *et al.*, 2007) has further found that the formation of a stable TatBC complex (430 kDa)

does not require TatA, in direct opposition to previous studies (Mangels *et al.*, 2005), and even that TatB and TatC can form homomultimeric complexes of 80 kDa and 220 kDa respectively. The authors go on to infer that the larger TatC complex would most likely form the core of a TatBC complex with TatB units bound peripherally unlike their previous claims from cysteine-scanning mutagenesis and crosslinking studies (Lee *et al.*, 2006). A separate study also concludes that TatC could act as the complex core due to the 250 kDa TatC complex they purified. However, the authors also observed a striking ladder of TatB (100-800 kDa), leading them to conclude that the formation of the TatBC complex acts to suppress such a potential multimerisation (Behrendt *et al.*, 2007). TatB and TatC have also been observed to homo-oligomerise by a number of different techniques including cross-linking and molecular fluorescence (De Leeuw *et al.*, 2001; Kostecki *et al.*, 2010; Lee *et al.*, 2006; Maldonado *et al.*, 2011). This interaction involves both the TM span and APH of TatB.

Single-particle EM was used for the first visual work on the purification and comparison of TatABC complexes in three bacteria: *E.coli, A.tumefaciens* and *S. typhimurium* (Oates *et al.*, 2003). These data showed functional and structural conservation of the TatABC complex across the species and revealed asymmetric complexes of varying size with dimensions of 9-13 x 10 nm. The ring was made of five to seven ~ 40 Å wide stain excluding densities surrounding a central pool of stain. Although these potential channels appeared too small to accommodate a fully folded protein of up to 100 kDa (50 Å in diameter), the correlation between particle size and the number of ring densities suggested a flexible modular organisation of the complex. In such a case more TatA subunits or several TatABC complexes would be needed to create a full sized channel. This was followed by a biochemical study into complex organisation which determined that TatABC exists as a single 370kDa complex made up of six to seven copies of TatBC with a similar number of TatA subunits (Oates *et al.*, 2005).

A more recent EM-based investigation into the TatBC complex has provided direct evidence that this complex represents the substrate binding site and how this binding occurs (Tarry *et al.*, 2009). Overproduction of TatBC and the SufI substrate in the absence of TatA led to the formation of clear TatBC-SufI complexes with strictly one or two substrate molecules bound peripherally to TatBC. Interestingly when two

substrate molecules are seen to bind they do so only at adjacent peripheral positions around TatBC, implying some structural or functional disparity between TatC protomers or negative cooperativity of substrate binding. The TatBC complex appears to lack the lidded channel shown for TatA and appears more roughly spherical with a small central cavity that does not alter size or shape upon substrate binding. A comparison with unliganded TatBC also revealed a slight reduction in the TatBC complex diameter upon substrate binding (~12 nm reduced to ~10 nm).

Although earlier studies suggested that TatA formed discrete homo-oligomeric complexes of 460 kDa (Porcelli *et al.*, 2002), more recent investigations have revealed complexes ranging from ~ 100 – 700 kDa (Barrett *et al.*, 2005; Gohlke *et al.*, 2005; McDevitt *et al.*, 2006; Oates *et al.*, 2005).

TatA complexes have been seen to form a defining ladder of bands ranging from 100-500 kDa in size by Blue Native-polyacrylamide gel electrophoresis (BN-PAGE). These complexes differed in size by ~ 34 kDa supporting a modular formation by units of three to four TatA molecules.

These striking TatA complexes were the target of the first investigation into the 3D structure of Tat complexes (Gohlke *et al.*, 2005). This subunit was shown to form channel-containing complexes of variable diameter leading to the proposal of the 'size-fitting pore' model. Using single particle EM in negative stain and a random conical tilt (RCT) reconstruction strategy a number of electron density maps were produced showing a channel with a lid present on one side; the results of protease accessibility studies of the TatA complex in a membrane environment suggest this lid to be exposed to the cytoplasmic face (Porcelli *et al.*, 2002). These models ranged from 85-130 Å wide and 45-55 Å high, comparable to the thickness of a membrane bilayer. At 30-70 Å wide the size range of the observed central pore matches well with the diameters of known *E.coli* Tat substrates (Berks, Ben C. *et al.*, 2000; Sargent *et al.*, 2002) leading to the dominant but highly contested theory that TatA complexes represent the translocation channel. A rough representation of the gross structure of these models is shown below in Figure 1.4.1.

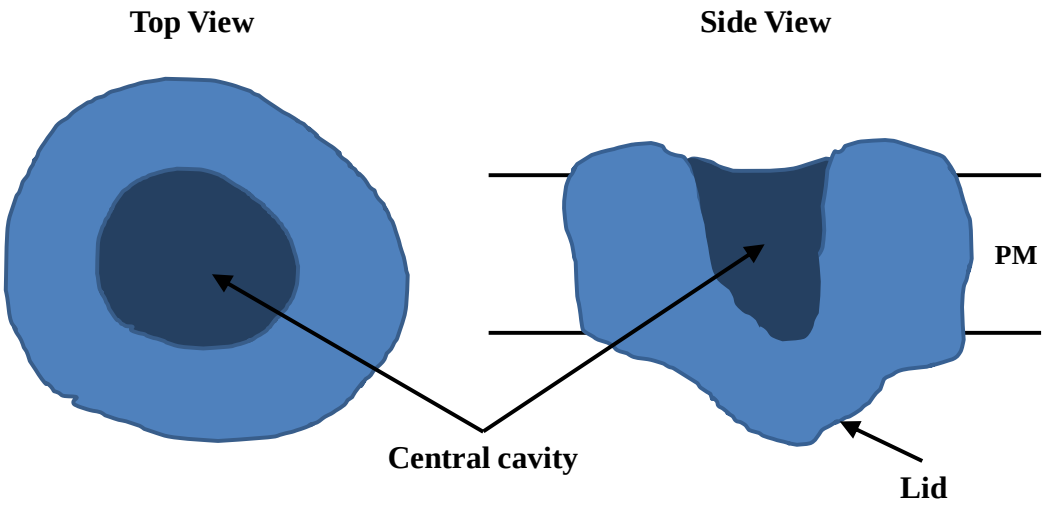



**Figure 1.4.1. Approximation of *E. coli* TatA complex gross structure**
A rough approximation of the TatA complex gross structure as revealed by single-particle EM and random conical tilt analysis (Gohlke *et al.*, 2005). All models consisted of a ring-shaped structure with a central cavity occluded by electron density on one side. The width of the ring was deemed sufficient to span the plasma membrane (PM).

This size-fitting pore model has been called into question by the purification and subsequent single-particle analysis of TatE complexes (Baglieri *et al.*, 2012). In the presence of native levels of TatB and TatC, TatE was shown to support efficient translocation of Tat substrates. However, unlike TatA, a distinctive ladder of BN-PAGE bands was not seen and although TatE was observed to form pore-like structures these were significantly smaller and more homogeneous in size than those of TatA. The lack of an observed size range in the TatA complexes of Gram-positive bacteria (see section below) also challenges this concept.
A more recent investigation into *E. coli* Tat complexes *in vivo* was unable to identify multimerisation of TatA in the absence of TatB and TatC using single molecule fluorescence (Leake *et al.*, 2008).

Together these data suggest that the fully formed translocon has yet to be isolated. The currently accepted model of a transiently coalescing unit suggests that the purification of such a complex would require the translocation event to be stalled in some way.

### 1.4.2. Tat complexes in Gram-positive bacteria

Such structural investigations into the Tat systems of Gram-positive bacteria, namely those of the model organism *Bacillus subtilis*, have been notably sparser. This may well be due to the marked variation between the Tat systems of Gram-positives in comparison to the relatively uniform system seen in Gram-negatives.

With the exception of the actinomycetes (Schaerlaekens *et al.*, 2001) and *Mycobacterium tuberculosis* (Dilks et al., 2003) these organisms generally contain only *tatAC* genes (Jongbloed *et al.*, 2004; Pop *et al.*, 2002) indicating a fundamental difference between these systems and those of Gram-negative bacteria. To date *B. subtilis* has been shown to contain two such AC-type Tat systems with differing substrate specificities (TatAdCd and TatAyCy) and a lone TatA (TatAc) (Jongbloed *et al.*, 2004), the function of which is not known. TatAdCd has been shown to form a ~ 230 kDa complex and a separate, discrete ~ 270 kDa complex of TatAd (Barnett *et al.*, 2008). The TatAyCy system has been shown to form similar discrete complexes of ~ 200 kDa for both TatAy and TatAyCy (Barnett *et al.*, 2009).

Such stable A-type complexes are markedly different from the size range seen for *E. coli* TatA and suggest a potentially different translocation mechanism in Gram-positive bacteria. The proposed bifunctional nature of TatAd (fulfilling TatA and TatB roles) (Jongbloed *et al.*, 2004; Jongbloed *et al.*, 2006) was demonstrated by its ability to complement *ΔtatA* and *ΔtatAE* mutants in *E. coli* (Barnett *et al.*, 2008). Furthermore, the TatAdCd system was shown to be capable of translocating a wide range of native *E. coli* substrates utilising complexes that do not present a large size range. The TatAyCy system does not appear as flexible as it is unable to transport the large *E. coli* Tat substrate TorA (Barnett *et al.*, 2009). As such these systems suggest that a discrete, but apparently flexible, translocon is sufficient to accept folded substrates of differing size.

Some studies have suggested that a proportion of TatAd may be present in the cytoplasm acting as a cycling receptor subunit, shuttling substrate to the cell membrane (Pop *et al.*, 2003; Westermann *et al.*, 2006). In one of these TatAd was analysed using negative stain and freeze-fracture EM (Westermann *et al.*, 2006). TatAd complexes of 150-250 kDa were purified in the absence of detergent by sucrose density gradient centrifugation and the electron micrographs revealed

particles that were heterogeneous in both size and shape ranging from 12 nm-100 nm in diameter. Freeze-fracture analysis of proteoliposomes revealed membrane-integrated TatAd complexes in the absence of TatCd, and immunogold labelled substrate prePhoD was used to show co-localisation to these complexes. From this it was inferred that TatAd represents the substrate recognition complex and is actively involved in the formation of the translocon relegating the C-type subunit from its substrate binding role seen in Gram-negatives.

Gram-positive Tat systems appear different from those of Gram-negatives and much work needs to be done to bring our knowledge of these systems up to the broader (but still as disputed) level of understanding shown for TatABC systems.

## 1.5. Tat translocation mechanisms

Despite many years of work the mechanism of Tat mediated translocation is still poorly understood and remains one of the most speculative and contested areas of the field. The ability to translocate fully-folded proteins without compromising the ionic seal of the membrane presents a particularly challenging and fascinating mechanistic problem, and there are several schools of thought on how this may be solved.

Translocation via the Tat system does not have a nucleoside triphosphate requirement, a unique feature amongst the mainstream protein transporters. However, transport is still dependent on the proton motive force (PMF), consisting of the $\Delta$pH and $\Delta\psi$. This feature was initially demonstrated in plants (Cline *et al.*, 1992; Klösgen *et al.*, 1992; Mould *et al.*, 1991) and later in *E. coli* (Alami *et al.*, 2003; Alami *et al.*, 2002; Yahr and Wickner, 2001). This requirement for the PMF appears to be limited to the translocation event itself and does not extent to the initial targeting of substrates to TatBC. A more recent *in vitro* investigation has indicated that two $\Delta\psi$-dependent steps are present during the translocation of *E. coli* Tat substrate SufI (Bageshwar and Musser, 2007).

A great deal of evidence using various experimental techniques points to the Tat(A)BC complex as the site of initial signal-peptide recognition (Alami *et al.*, 2003; Alami *et al.*, 2002; Cline and Mori, 2001; McDevitt *et al.*, 2006). Cross-

linking studies have identified a number of interactions with TatC which are then followed by more extensive contacts to TatB.

Recent evidence has suggested that the signal peptide can contact the TM span of TatA prior to translocation leading the authors to suggest a close proximity between TatA and TatC. Interestingly the interaction between TatA and the signal peptide was only seen in the presence of a PMF (Fröbel, J. *et al.*, 2011), supporting the idea of a transiently coalescing translocase

In all current models TatA-type proteins are presumed to play a major role. This 'size-fitting pore' model is based heavily around the observation of pore-like structures of variable diameter in *E. coli* by single-particle EM (Gohlke *et al.*, 2005). It has been observed that TatBC is required to induce oligermerisation of TatA (Leake *et al.*, 2008). A model of the proposed mechanism of Tat translocation is shown in Figure 1.5.1.

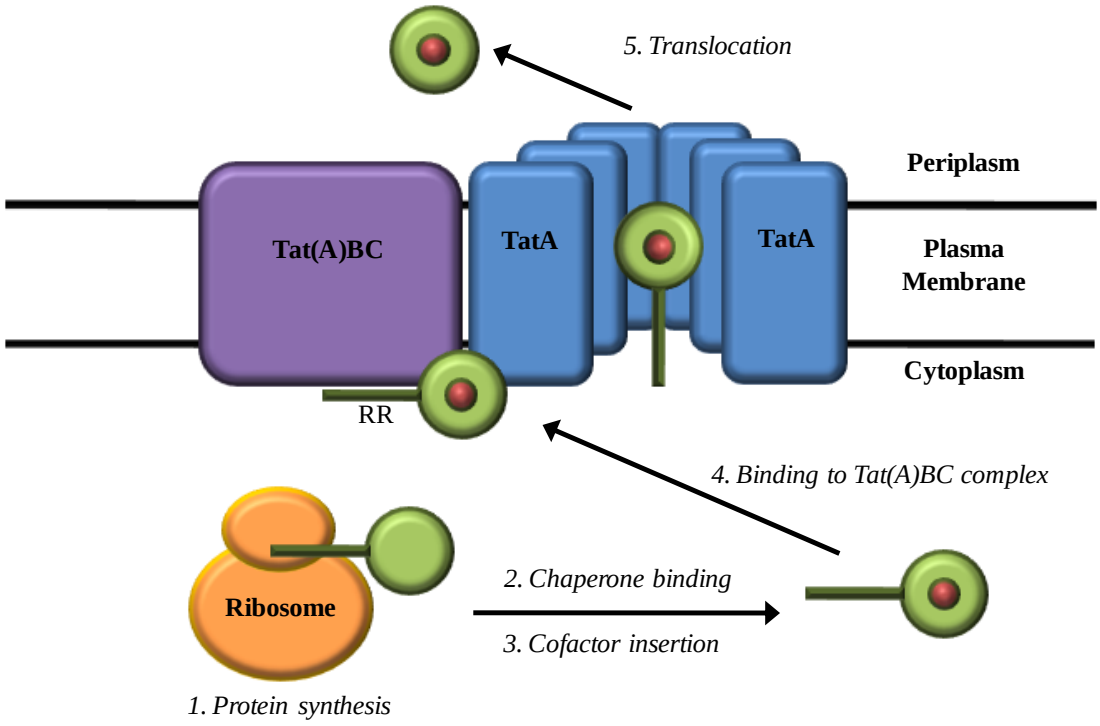



**Figure 1.5.1. Tat translocation based on the current accepted mechanism**
After folding the substrate associates with the Tat(A)BC substrate binding unit triggering a transient coalescence of multiple TatA units to form a channel across the membrane through which the folded substrate can pass.

An alternative model proposed by Brüser and Sanders suggests that a structural pore is not formed at all (Brüser and Sanders, 2003). In this model the accumulation of TatA-type monomers around the Tat(A)BC complex would induce membrane destabilisation to an extent where the Tat substrate could diffuse through. This method would circumvent the need for size fitting pores and potentially unifies the structural observations in both Gram-negative and Gram-positive bacteria.

Upon reviewing the field as it currently stands, the lack of direct structural data on Tat complexes stands out as the most immediate and pressing challenge. This information is vital for elucidating the translocation mechanism and how such a system can preserve membrane integrity.

## 1.6. Biological transmission electron microscopy (TEM)

Electron microscopy (EM) is one of the three major methods used for obtaining a high resolution structure of biomolecules. Unlike X-ray crystallography and nuclear magnetic resonance (NMR), EM techniques can be used to visualise a huge size range of structures; from whole cell sections and tissue samples down to individual proteins and approaching atomic resolution. Via several methods high resolution 3D electron-density maps can be produced using EM and downstream image processing techniques. These approaches are especially relevant for integral membrane proteins and multimeric complexes which do not crystallise.

### 1.6.1. Principles of EM image formation

In transmission electron microscopy (TEM) images of very thin objects are formed by passing high energy electrons through them. As the electrons have a much shorter wavelength compared to light the maximal resolution is much higher than a light microscope. The beam of electrons is directed and focused using electromagnetic lenses to pass through the specimen, and an image is formed directly on a fluorescent screen by scattered electrons; whilst passing through the specimen these electrons can interact with the matter which affects their scattering angle. If no energy is

transferred between the sample and the electron, it is elastically scattered and contributes to the desired high-resolution information. However, if energy is transferred the electron is inelastically scattered and contributes to the unwanted background noise in the image. As the electrons can also be scattered by air, the column of the microscope is maintained in a high vacuum.

In the most commonly used imaging mode (Bright-field EM), contrast in the image results from the interference of the 'unscattered' electron wave with the (elastically) 'scattered' wave. The scattered wave is out of phase with the unscattered wave, producing the phase contrast. Additionally, amplitude contrast is produced by the loss of electrons from the scattered wave, due to scattering outside the aperture or by inelastic scattering. Overall the image formation can be described by the action of the contrast transfer function (CTF) which modifies the object contrast to produce the image with a succession of rotationally symmetric positive and negative zones, as explained in Equations 1.6.1 and 1.6.2 and shown in Figure 1.6.1. Various technical limitations of the microscope, described by the compound envelope function ($E$(k)), see Equation 1.6.3, dampen the CTF at higher spatial frequencies and in turn limit the resolution. Although the CTF generates contrast it has several degrading effects on the information content of the image. These can be described as a combination of low-pass (resolution-limiting) and high-pass filtration along with several reversals of contrast resulting in the loss of information over some resolution ranges. These degrading effects have to be corrected for later in order to restore the information content of the image, at the expense of a loss of contrast. This is in turn compensated for by averaging together many copies of similar images and vastly improving the signal to noise ratio.

$$I(\mathbf{r}) = \Phi(\mathbf{r}) \circ h(\mathbf{r})$$

**Equation 1.6.1. Formation of EM image contrast**
This describes in real space the relationship of the image ($I(\mathbf{r})$) to the projected object potential ($\Phi(\mathbf{r})$) modified by a convolution operation. $I(\mathbf{r})$ = Observed image contrast, $\Phi(\mathbf{r})$ = Projected object potential, $h(\mathbf{r})$ = The point spread function, $\circ$ = Represents multiple convolutions

$$F(\mathbf{k}) = O(\mathbf{k})\, A(\mathbf{k}) \sin \gamma(\mathbf{k})$$

**Equation 1.6.2. The impact of the CTF in EM image contrast**
In this way the CTF (sin $\gamma(\mathbf{k})$) is used to provide a linear relationship between the Fourier transforms of the image ($I(\mathbf{r})$) and the projected object ($\Phi(\mathbf{r})$) along with the aperture function ($A(\mathbf{k})$). $F(\mathbf{k})$ = FT of $I(\mathrm{r})$, $O(\mathbf{k})$ = FT of $\Phi(\mathrm{r})$, $A(\mathbf{k})$ = Aperture function, sin $\gamma(\mathbf{k})$ = CTF

$$E(k) = E_i(k)\, E_e(k)$$

**Equation 1.6.3. The compound envelope function**
This is an estimation of the impact of technical limitations on the maximal attainable resolution. These factors are due to how the electron beam is generated. In instruments using a Field Emission Gun (FEG) electron source both envelopes are small so higher resolutions can be attained. $E(k)$ = the compound envelope function, $E_i(k)$ = the envelope function due to partially coherent illumination, $E_e(k)$ = the envelope function due to energy spread

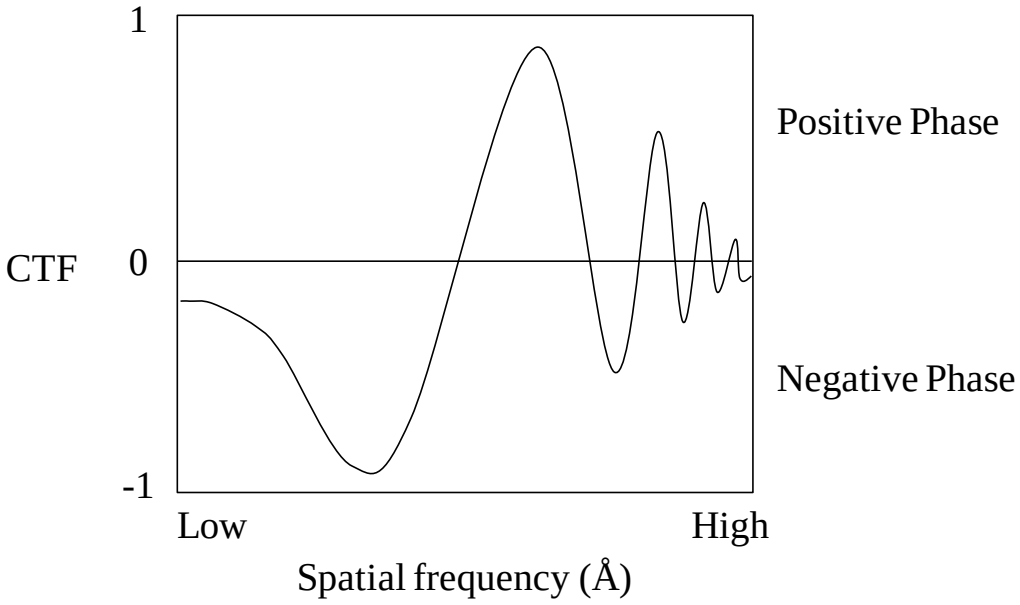



**Figure 1.6.1. The contrast transfer function (CTF)**
The contrast transfer function produces an alternating series of positive and negative contrast bands. These decrease in amplitude but increase in oscillation frequency when moving towards higher resolutions (spatial frequency). The position of the bands over the resolution range of the image is dependent on the defocus of the image. By estimating the image defocus the relative position of the CTF can be fitted. From this, the position of the zeros can be estimated (the point where the phase alternates). When correcting the CTF by phase flipping, the negative zones of the CTF are flipped to the positive, resulting in a continuous additive signal with no oscillating bands, thereby restoring the information content of the image.

Overall the contrast of biomolecules within these images is very low as their constituent atoms are not very dense (have low atomic numbers) and therefore do not scatter electrons very well. By introducing defocus to the image the relative contrast of low resolution features, such as the outline of the particle, can be improved at the expense of high-resolution internal features. In this way the selected defocus at which the image is taken is a trade-off between these factors.

### 1.6.2. Negative stain

This method produces high contrast in the image due to the addition of heavy metal stains that scatter electrons very well. However the structural information is limited to the shape of the molecule as it appears in the 2D projection, which can be distorted during drying of the stain or by partial staining of the object. The sections of the object that stick out from the stain cannot be seen and this can be compounded by variations in the thickness of the stain across the grid. Despite these drawbacks negative stain does have some advantages over more detailed cryo-EM methods. The sample is effectively fixed during staining and is much more resistant to electron beam induced degradation; the contrast is significantly higher than in cryo-EM making the particle boundaries much easier to distinguish; and the technique can be used for high through-put screening experiments. As such negative staining with a wide variety of heavy metals remains a vital first step in a 3D EM investigation and is especially useful for unknown or difficult samples.

### 1.6.3. Cryo-EM

Using cryo-EM much higher resolutions can be obtained at the expense of a loss of contrast and more technically demanding experimental conditions. The sample is snap-frozen in a very thin layer of vitreous ice using liquid ethane, producing a pseudo-aqueous state. The loss of contrast is due to the absence of stain and the ice itself, which scatters electrons at a very similar angle to protein. The ice and protein are quickly damaged under the electron beam and precautions must be taken to use the lowest electron dose for the least amount of time.

## 1.7. Image processing

Micrograph images can be acquired using electron imaging film or digitally using a charge-coupled device (CCD) camera. Micrographs taken by TEM represent 2D density projections of the imaged objects; as such they contain information on the internal structure of the 3D specimen. When imaging a protein sample, multiple copies of the same structure are spread out over the surface of the grid, lying in a range of different orientations. These particles are windowed out of the micrographs into boxes of a set size to provide the processing data-set. Through the careful sorting of averaged images anomalous particles representing aggregates, contaminants or broken complexes can be removed. A single particle analysis technique can then be used to computationally sort these different views and align them in 3D space to reconstruct a model of the protein. In some cases the number of views present is very limited due to certain orientations of the protein being preferred. This can relate to the shape of the complex or to the interaction of a specific region of the macromolecule with the grid.

### 1.7.1. Random conical tilt (RCT)

If the sample presents a small number of preferred orientations then a random conical tilt (Radermacher, Michael, 1988; Radermacher, M. *et al.*, 1987) reconstruction method is required to provide the necessary range of views of the structure in 3D space. In this method the grid is first imaged while the stage is tilted by a fixed angle (normally 45°) so particles showing 'identical' views now appear as a conical projection series with random azimuth, relating to a random rotation relative to the untilted particle position. An image of the same area is then taken with the stage untilted. The untilted particles can be aligned using translational and rotational cross-correlation functions to inform the positioning of their tilted counterparts along the azimuth. The orientational relationship between the tilted particles using this method is detailed below in Figure 1.6.2.

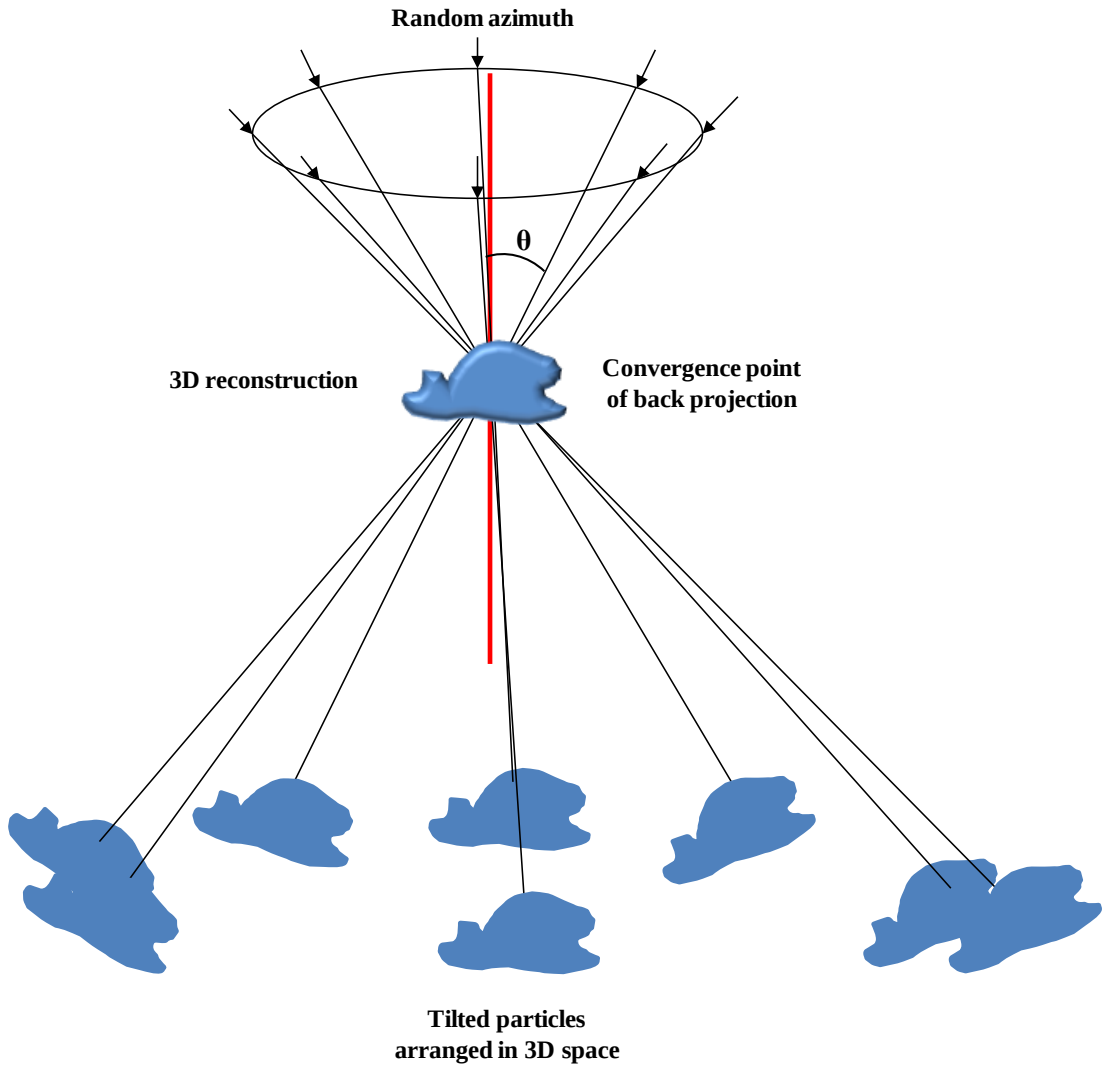



**Figure 1.6.2. 3D reconstruction using the Random Conical Tilt method**
The position of each tilted particle along the random azimuth is determined by the rotational alignment of the untilted particles. Once arranged in 3D space the tilted particles are then back projected and brought to the convergence point to produce a 3D model of the original object. The red line represents the untilted (0°) z-axis, θ (Theta) is the tilt angle of the microscope stage.

### 1.7.2. Multivariate statistical analysis and classification

Simply aligning and averaging all the particles in the data-set together will not produce a meaningful structure. The particles must be sorted into classes by similarity to produce meaningful averages with low intra-class variance. A wide number of factors introduce variance within the data-set, some are dependent on the sample (e.g. orientations, conformers, substrate bound complexes, degradation products, contaminants) and some are dependent on the imaging method (e.g. variable staining, variable contrast, localised tilts of the grid, electron beam induced damage, S/N ratio).

To sort through these factors, multivariate statistical analysis (MSA) methods can be applied to investigate the impact of different features of variance on the data set as a whole. Correspondence analysis (CA) is one such method, applied to categorical data, which can display the relationship between two factors in factorial space. The clustering of the images based on these factors in 2D space can be used to determine the significance of each factor of variance. Once significant variance factors have been identified the data-set can be classified into groups of particles based on these factors. Hierarchical ascendant classification (HAC) is an unsupervised method of clustering data into classes based on similarity. The constituents of the data-set are successively merged into larger and larger groups or clusters based on their distance in factor space. A range of merging rules can be applied to control how the data clusters, with each merger resulting in an index of similarity, used to construct a tree of hierarchical relationships referred to as a dendrogram. The commonly used Ward's clustering criterion (Ward Jr, 1963) uses minimal added intraclass variance as the merging rule.

### 1.7.3. Three-dimensional reconstruction and refinement

Once the image data-set has been successfully divided into groups of similar particles, 3D electron density maps can finally be generated using a range of different 3D reconstruction methods. These methods vary in both their accuracy and speed of structure generation based on the computational algorithms used. More accurate methods require additional information or 'restraints' to be provided by the user to limit how the images can align in 3D space.

Once an initial 3D model is generated the 3D alignment, and hence the resolution, can be improved using an iterative projection matching method (Penczek *et al.*, 1994). This involves generating a range of 2D projections of the initial model covering equidistant regions of 3D space. The particles are then compared to the full range of these projections and aligned to the most similar, before a refined model is generated. In this way each particle is allowed to find a 'better fit' in 3D space.

## 1.8. Aims and objectives

The overall aim of this study was to initiate and develop a structural investigation into bacterial Tat protein complexes by single particle electron microscopy.

The investigation centred on the components of the *Escherichia coli* and *Bacillus subtilis* Tat pathways as the Gram negative and Gram positive model organisms respectively. These were selected as the vast majority of previous work in this field focuses on these organisms.

Key aims and questions addressed in this work include:

Do Tat complexes other than *E. coli* TatA form ring shaped particles?

Do these particles vary in size, if so by how much? Could they represent size fitting pores?

How does detergent concentration affect the observed particle populations?

How is the structure of Tat complexes affected by loss of function mutations?

Produce 3D structural models of purified Tat complexes.

In order to answer these questions two levels of image processing procedures had to be implemented:

Develop and implement an in-house image processing strategy for assessing and separating different size particles.

Develop and implement an in-house processing strategy to generate 3D structural models of Tat complexes using a random conical tilt reconstruction strategy.

# *Chapter 2*

# *Materials and Methods*

## 2.1. Suppliers of reagents, chemicals and materials

All reagents and materials were obtained from the companies indicated below.

**Agar Scientific (UK):** formvar/carbon coated copper grids (300 mesh).

**GE healthcare (UK):** Superdex 200 10/300 GL pre-packed gel filtration column; ECL™ detection reagents; Native protein markers for gel electrophoresis and gel filtration chromatography; Hybond TM-P PVDF membrane..

**Takara Bioscience (UK):** Talon™ affinity chromatography resin.

**Calbiochem (Germany):** *n*-Dodecyl-*ß*-D-maltoside (DDM)

**Expedeon (UK):** Instant Blue™ ready-to-use Coomassie® stain.

**Fermentas (UK):** Bradford Reagent, ready-to-use

**Fisher Scientific (UK):** Acetic acid (glacial); Acetone; Ethanol; Formaldehyde; Glycerol; Glycine; Hydrochloric acid; Methanol; Sodium chloride; sodium dodecyl sulphate; Sucrose, and Tris.

**Fuji (Japan):** Super RX film.

**IBA (Germany):** Buffer E; Strep-tactin™ HRP-conjugate, and Strep-tactin™ affinity chromatography resin.

**Invitrogen (USA):** Anti-His (C-terminal) antibody.

**New England Biolabs (USA):** Pre-stained broad range protein markers.

**Premier International Foods (UK):** Marvel milk powder.

**Promega (UK)** Anti-rabbit-HRP and anti-mouse-HRP conjugates

**Roche applied science (UK):** Complete™ protease inhibitor cocktail tablets.

**Sigma-Aldrich (UK):** Ampicillin; L-arabinose; Avidin; Bis-Tris; ßmercaptoethanol; Bovine serum albumin; Nonaethyleneglycol mono-n-Dodecyl Ether ($C_{12}E_9$); Imidazole; Lysozyme; Silver nitrate; Sodium carbonate; Sodium thiosulphate; TEMED; Tricine; and Trizma.

**VWR (UK):** 40% Acrylamide solution; Ammonium persulphate; Bromophenol blue; EDTA; Glucose; Magnesium chloride; Sodium hydroxide pellets; Trichloroacetic acid and Tween 20.

**Whatman (UK):** 3mm filter paper.

## 2.2. Bacterial cultures

### 2.2.1. *E. coli* strains and plasmids used

The *E. coli* strain and bacterial plasmids used in this work are presented below in Table 2.1.1.

| Strain | Properties | Reference |
|---|---|---|
| MC4100 | *F-, araD139, Δ*(*ara-leu*)7696; *Δ*(*lac*)X74, *galU*, *galK*, *hsdR*2, *mcrA*, *mcrB*1, *rspL* | Casadaban and Cohen, 1980 |
| *ΔtatABCDE* | MC4100 *ΔtatABCDE*, *Ara*$^r$ | Sargent *et al.*, 1988 |
| *ΔtatA/E* | MC4100 *ΔtatA/E*, *Ara*$^r$ | Sargent *et al.*, 1988 |
| pBAdh | pBAD24 + *B. subtilis tatAd-his* | Barnett *et al.*, 2008 |
| pBAD-Es | pBAD24 + *E. coli tatE-strep* | Sargent *et al.*, 1988 |

**Table 2.1.1. Bacterial strains and plasmids used.**
All bacterial strains and plasmids used in this work along with relevant properties and the initial reference where the constructs were first used.

### 2.2.2. Antibiotic supplements

Ampicillin (100 µg/ml) was used in cultures containing the pBAD24 plasmid.

### 2.2.3. Storage of *E. coli* cells

Glycerol stocks of *E. coli* cultures were prepared for long-term storage by mixing 2 parts liquid culture with 1 part 50% (v/v) glycerol. Cultures were frozen on dry-ice and stored at -70 °C.

### 2.2.4. Cell culture and induction of plasmids

Cells were grown aerobically in Luria broth (LB) until they reached mid-exponential growth phase ($OD_{600}$ ~ 0.6) before induction of the pBAD24 plasmid by the addition of 500 µM arabinose to the growth medium. Cells were then left to grow for a further 2-3 hours.

## 2.3. Fractionation of *E. coli* cells

Cell samples were separated into periplasmic, cytoplasmic and membrane fractions using the well documented lysozyme/cold osmotic shock method (Randall and Hardy, 1986). A 1.5 L *E. coli* culture was subjected to centrifugation to pellet out the cells, these were then resuspended in 72 ml ice cold disruption buffer (100 mM tris-acetate pH 8.2, 0.5 M sucrose + 1.5 dissolved Complete – EDTA protease inhibitor tablets). To this resuspension 3 ml of 2 mg/ml lysozyme were added and the samples were incubated on ice for 5 minutes. 72 ml of ice cold $dH_2O$ were then added along with 3 ml of 1.0 M $MgSO_4$, and the sample was gently mixed. The resulting spheroblasts were then pelleted by centrifugation at 4000 g for 15 minutes. The periplasmic fraction (supernatant) was discarded and the pellet was resuspended in 30 ml of sonication buffer (50 mM tris-acetate pH 8.2) before sonication at 10 microns using three 30 second pulses with 1 minute rest intervals. Having broken the cells membranes these were separated from the cytoplasmic fraction by ultracentrifugation at 70000 rpm for 35 minutes using a Beckman Tl100.3 rotor. The

isolated membranes were finally solubilised in 15 ml solubilisation buffer (20 mM tris-HCl pH 8.0, 50 mM NaCl and 1% (w/v) *n*-dodecyl-*β*-D-maltoside.

## 2.4. Protein chromatography

### 2.4.1. Affinity tags

For the purification and detection of proteins C-terminal tags were used. Either the 10 amino acid *Strep*-tag II (IBA, Germany) or a hexa-histidine tag were used where stated in the text, denoted by *strep* or s, and *his* or h, respectively.

### 2.4.2. Talon™ affinity column chromatography

10 ml of Talon™ affinity resin was equilibrated in the same volume of equilibration buffer (20 mM Tris-HCl pH 8.0, 400 mM NaCl, 5 mM Imidazole and 0.02% DDM) overnight with rotation at 4°C. The buffer was then removed by centrifugation and solubilised membranes, prepared as indicated in 2.2.3 were added and again the slurry was left to rotate overnight under the same conditions. This method was used to facilitate optimal binding of the protein to the entirety of the Talon™ resin. The mixture was then poured into a disposable column and left to settle for approximately 30 minutes before elution of the flow through. The column was subsequently washed three times with 2 column volumes of equilibration buffer. Bound protein was eluted from the column using 20 ml of equilibration buffer containing 150 mM imidazole.

### 2.4.3. Streptactin™ affinity column chromatography

Performed as per the Talon™ affinity column above, but using 4 ml of Streptactin™ resin. Additionally, samples were incubated with 1 µM avidin for 30 minutes prior to their addition to the column, to prevent the non-specific binding of biotinylated proteins. The equilibration buffer contained 20 mM Tris-HCl pH 8.0, 150 mM NaCl, and 0.02% DDM. Bound protein was eluted with 12 ml buffer E (20 mM Tris-HCl, 150 mM NaCl, 3mM desthiobiotin) plus 0.02% DDM.

### 2.4.4. Gel filtration chromatography

Size exclusion chromatography was used as a final clean-up procedure for the protein purification and as an analytical tool to estimate the mass of protein complexes.

An ÄKTA purifier FPLC system (GE healthcare) operating UNICORN® v.4.00 software was used in conjunction with Prepacked Tricorn™ GL columns (GE healthcare). A Superdex 200 10/300 GL column (separation range: 10 – 600 kDa globular protein) was used to provide optimum resolution over the target size range. The column was pre-equilibrated with at least 2 column volumes (CVs) of the selected GF buffer (20 mM Trizma-HCl, 150 mM NaCl) with or without detergent supplement. Sample separation was routinely performed at 0.5 ml/min and the elution volume was collected in 0.5 ml fractions. Between sample runs the column was rinsed with 2 CVs buffer and the 0.5 ml loading loop was manually flushed with the same buffer. The following cleaning-in-place (CIP) protocol was used routinely after daily usage to maintain column performance and resin purity: Wash with 2 CVs $dH_2O$, then 2CVs 0.5 M NaOH, then 2CVs $dH_2O$.

### 2.4.5. Estimation of protein complex mass

The gel filtration column was calibrated using protein standards of known molecular mass in order to prepare a standard curve for size estimation. Blue Dextran was used to estimate the column void volume (Vo); the volume at which all matter too large to enter the resin matrix is eluted. The protein standards and their elution volumes (Ve) are shown below in Table 2.4.5.

| Mol. Wt. (x10-3) | Mol. Wt. | Protein standard | Ve | Ve/Vo |
|---|---|---|---|---|
| 669 | 669000 | Thyroglobulin | 9.30 | 1.13 |
| 440 | 440000 | Ferritin | 10.92 | 1.32 |
| 158 | 158000 | Aldolase | 12.87 | 1.56 |
| 66 | 66000 | Albumin | 14.03 | 1.70 |
| | | | **Vo** | |
| 2000 | 2000000 | Blue Dextran | 8.25 | |

**Table 2.4.5. Calibration of the Superdex 200 gel filtration column**
Known molecular weights (Mol. Wt.) of soluble protein standards are presented in Daltons. The elution volumes (Ve) of each standard are shown, along with the void volume (Vo) corresponding to the elution volume of Blue Dextran. A linear plot of Ve/Vo is used to estimate the size of the unknown protein sample.

### 2.4.6. Densitometry of gel filtration elution fractions

2D densitometric analysis of Western blot bands was performed using AIDA (Advanced Image Data Analyzer) software v. 3.28.001.

## 2.5. Estimation of protein concentration

Estimation of protein concentration was performed using Bradford's reagent (Fermentas) as per the manufacturer's method: 20 µl of sample mixed with 1 ml room-temperature Bradford's reagent, assay readings taken at 595 nm after 5 minute incubation. Standard curves were prepared using Bovine serum albumin (BSA) dilutions.

## 2.6. Protein Electrophoresis

### 2.6.1. SDS poly-acrylamide gel electrophoresis (SDS-PAGE)

Gel electrophoresis was performed using the C.B.S vertical gel system and a uniform running buffer (25 mM Tris, 250 mM glycine and 0.1 % SDS). SDS-PAGE gels consisted of a 17.5% resolving gel (17.5% acrylogel, 375 mM Tris-HCl pH 8.8, 0.1% APS and 0.06% TEMED) and a 4% stacking gel (4 % acrylogel, 125 mM Tris-HCl pH 6.8, 0.1 % SDS, 0.6 % APS and 0.06 % TEMED). Protein samples were

prepared by diluting aliquots 1:1 with 2x loading dye (125 mM Tris-HCl pH 6.8, 20 % glycerol, 4 % SDS, 0.02 % bromophenol blue and 5 % ß-mercaptoethanol) and heating to 70°C for 5 minutes. Gels were left to run at 35 mA for ~ 2.5 hours or until desired separation was achieved.

## 2.7. Protein detection

### 2.7.1. Silver staining

After separating proteins by SDS-PAGE, gels were incubated for 15 minutes in fixer solution (50 % acetone, 1.25 % TCA and 0.015 % formaldehyde). The fixer was then removed and the gel washed repeatedly for 5 minutes with $dH_2O$, before incubation with 50 % acetone for 5 minutes. This was followed by a 1 minute incubation with a 0.02 % (w/v) sodium thiosulphate enhancer solution before additional washing in $dH_2O$. The gels were then stained for 8 minutes using a solution containing 0.25 % (w/v) silver nitrate and 0.4 % formaldehyde and washed with $dH_2O$ before inbuation in developer solution (0.2 mM sodium carbonate, 0.004 % (w/v) sodium thiosulphate and 0.015 % formaldehyde). Once the protein bands were sufficiently visible the staining process was stopped using 1 % acetic acid solution. The gel was finally washed in $dH_2O$ before imaging.

### 2.7.2. Coomassie staining

After SDS-PAGE gels were incubated with Instant Blue™ Coomassie stain for over 1 hour before removal and subsequent washing in $dH_2O$ before imaging.

### 2.7.3. Protein transfer to PVDF membranes

A semi-dry Western blotting apparatus (Sigma, UK) was used to transfer proteins from SDS-PAGE gels onto PVDF membranes. After wetting the electrodes of the blotter two sheets of Whatman paper soaked in the transfer buffer were placed on the positive electrode. A methanol soaked PVDF membrane was then layered on top of the Whatman papers, followed by the gel. Lastly, two more pieces of buffer soaked

Whatman paper were placed over the gel before closing the blotter. A uniform Towbin transfer buffer was used (25 mM Tris, 192 mM glycine and 20% methanol) (Towbin *et al*., 1979), and proteins were transferred at 200 mA over ~ 2 hours or at 20 mA over-night.

### 2.7.4. Immunoblotting

Following transfer of the proteins to PVDF membrane these were blocked to reduce non-specific binding of antibodies using either 5% (w/v) dried milk powder or 2% (w/v) bovine serum albumin (BSA) (for anti-*strep* II) dissolved in PBS-T. The membranes were left in blocking solution normally overnight, but at least more than 1 hour. After this the membranes were rinsed in PBS-T before the addition of 20 ml of the relevant primary antibody, diluted as shown in Table 2.7.4, for 1 hour. For *strep* II tagged proteins, the membranes were incubated in 20 ml of biotin blocking buffer (at 1:10000 dilution) for 10 minutes before the addition of the anti-*strep* II HRP conjugate for 2 hours. For the other primary antibodies, following repeated washing in PBS-T over 1 hour, the corresponding secondary HRP conjugate antibody was applied to the membrane; this was then incubated for 1 hour. Finally all membranes were washed repeatedly in PBS-T over at least 1 hour. Detection of the proteins was achieved using the EZ-ECL detection kit and X-ray films were developed using an AGFA Curix 60 automatic developer, as per the manufacturer's instructions in each case. Antibodies used are listed below in Table 2.7.4.

| Antibody | Dilution | Source |
|---|---|---|
| Mouse anti-*his* (C-terminal) | 3:20000 | Invitrogen, USA |
| Mouse anti-*strep* II HRP conjugate | 1:10000 | IBA, Germany |
| Rabbit anti-TatAy | 1:10000 | Jan Maarten van Dijl (University of Groningen) |
| Anti-rabbit IgG HRP conjugate | 1:10000 | Promega, USA |
| Anti-mouse IgG HRP conjugate | 1:10000 | Promega, USA |

**Table 2.7.4. Antibodies in this investigation**
List of all antibodies used in this work along with the working dilution and original source of acquisition.

## 2.8. Electron microscopy gridding methods

### 2.8.1. Sample preparation for EM in negative stain

20 µl aliquots were taken from the relevant gel filtration elution fraction. These were stored at -20°C prior to use. Any dilutions were performed using the standard gel filtration buffer solution (20 mM Trizma-HCl, 150 mM NaCl) with or without detergent where stated. All solutions were filtered by hand through a 0.2 µm filter before use in dilution or washing steps.

### 2.8.2. Grids and stains used

**Agar Scientific (UK):** Formvar/carbon coated copper grids (300 mesh); lacey carbon grids; Uranyl acetate 2% or 1% (w/v) in $dH_2O$ (passed through 0.22 µm filter prior to use).

### 2.8.3. Glow discharge

Glow discharge was performed with an EMtech K100x unit, running at $3x10^{-1}$ mbar and 25ma for 15 seconds of negative discharge in air.

### 2.8.4. Touching drop gridding method

The standard gridding method is given here; any modifications to this are presented later in the text where relevant:

4 µl of sample was applied to a grid within 10 minutes of glow-discharging and incubated on the grid for 1 minute before washing by touching the sample to an 8 µl drop of GF buffer for ~ 10 seconds with agitation. The grid was then blotted before quickly touching to a second drop of buffer and repeating this process. Next the grid was washed for ~ 10 seconds with an 8 µl drop of stain before blotting and staining with a second drop for 20 seconds. The grid was carefully blotted a final time before being left to dry. Adapted from a previously published method (Rubinstein, 2007)

using a Parafilm® drop support and Whatman® paper for blotting, grids were left to air dry.

### 2.8.5. Cryo-plunging

For cryo-imaging lacey carbon grids were used. These are grids with a large number of irregular sized holes present in the carbon film where an ultra-thin layer of vitreous ice will form in which the particles of interest are imbedded. ~7 µl of sample was applied to a lacey carbon grid for 1 minute before blotting both sides simultaneously for 5 seconds and plunging the grid into liquid ethane following standard cryo-plunging methodology. An in-house machined pneumatic plunge system was used to achieve this.

## 2.9. Transmission electron microscopy (TEM)

### 2.9.1. JEOL 2010F TEM

200 kV JEOL transmission electron microscope with a field emission gun (FEG) electron source, operating a Gatan Ultrascan™ 4000 CCD camera with a pixel size of 14 µm.

### 2.9.2. Imaging software

The microscope was controlled via the FasTEM software suite (ver.3) and the CCD camera was operated using the Digital Micrograph package (ver.1.81.78).

### 2.9.3. Microscope alignment protocol

1. Switch VALVE to open position.
2. Centre the beam using the condenser aperture controls.
3. Check spot size 1, centre with SHIFT X+Y.
4. Check spot size 5, centre with IMAGE SHIFT DEF X+Y.
5. Check spot size 1 again, centre with GUN SHIFT X+Y.
6. Switch on ANODE WOBB, align roughly using GUN DEF so cross-over is central (not a swinging hourglass shape).

7. Use COND STIG; reduce the spot size as much as possible using DEF X+Y.
8. Finish alignment of ANODE WOBB (see point 6).
9. Turn on HT and BRIGHT TILT; use DEF X+Y to centralise the pulse.
10. Repeat steps 3-5.
11. Close VALVE and complete grid insertion.
12. Use IMAGE WOBB and set the Z-height.
13. Repeat step 9.
14. Check ANODE WOBB again.
15. Insert CCD camera and switch on OBJ STIG; use DEF X+Y on live FFT view to produce a perfectly circular pattern of rings.

### 2.9.4. Random Conical Tilt imaging

An in-house machined grid mounting blade was used to allow the pole-piece to rotate more than 40° within the JEOL 2010F. The stage was manually tilted (TX) to either 45° or 50° prior to capturing the first image of the tilt pair. Magnification was then reduced and the beam spread increased whilst the stage tilt was reset. During this process the imaging position was maintained by constant manual adjustment with the track-ball. The magnification and beam-spread was then re-set to image capture conditions and the untilted image was acquired. All images were taken at a range of defoci.

## 2.10. An EM image processing protocol for size-based particle classification

### 2.10.1. Introduction

This section focuses on the image processing techniques adopted in these structural investigations into Tat complexes. Here a schematic overview of the implemented procedures is provided, followed by a detailed step-by-step protocol of the strategy. The protocol relates to the computer scripting documented in the appendix chapter for reference purposes. For a detailed overview of the theory behind these techniques see: (Frank, 2006).

The majority of the computational image processing procedures were performed using SPIDER (System for Processing Image Data from Electron microscopy and Related fields) operations (Frank *et al.*, 1981). These operations were run

consecutively as batch files (scripts) organised and executed using the SPIDER Reconstruction Engine (SPIRE ver. 1.5.3) (Baxter *et al.*, 2007) which provides an adaptable graphical user interface (GUI). All of the SPIDER batch files presented in this study were modified from, or at least informed by, those provided in the SPIRE tutorial package (Shaikh *et al.*, 2008). The SPIRE tutorial was accessed online at: www.wadswoth.org/spider_doc/spire/doc/download.html where it is provided as an open source distribution package.

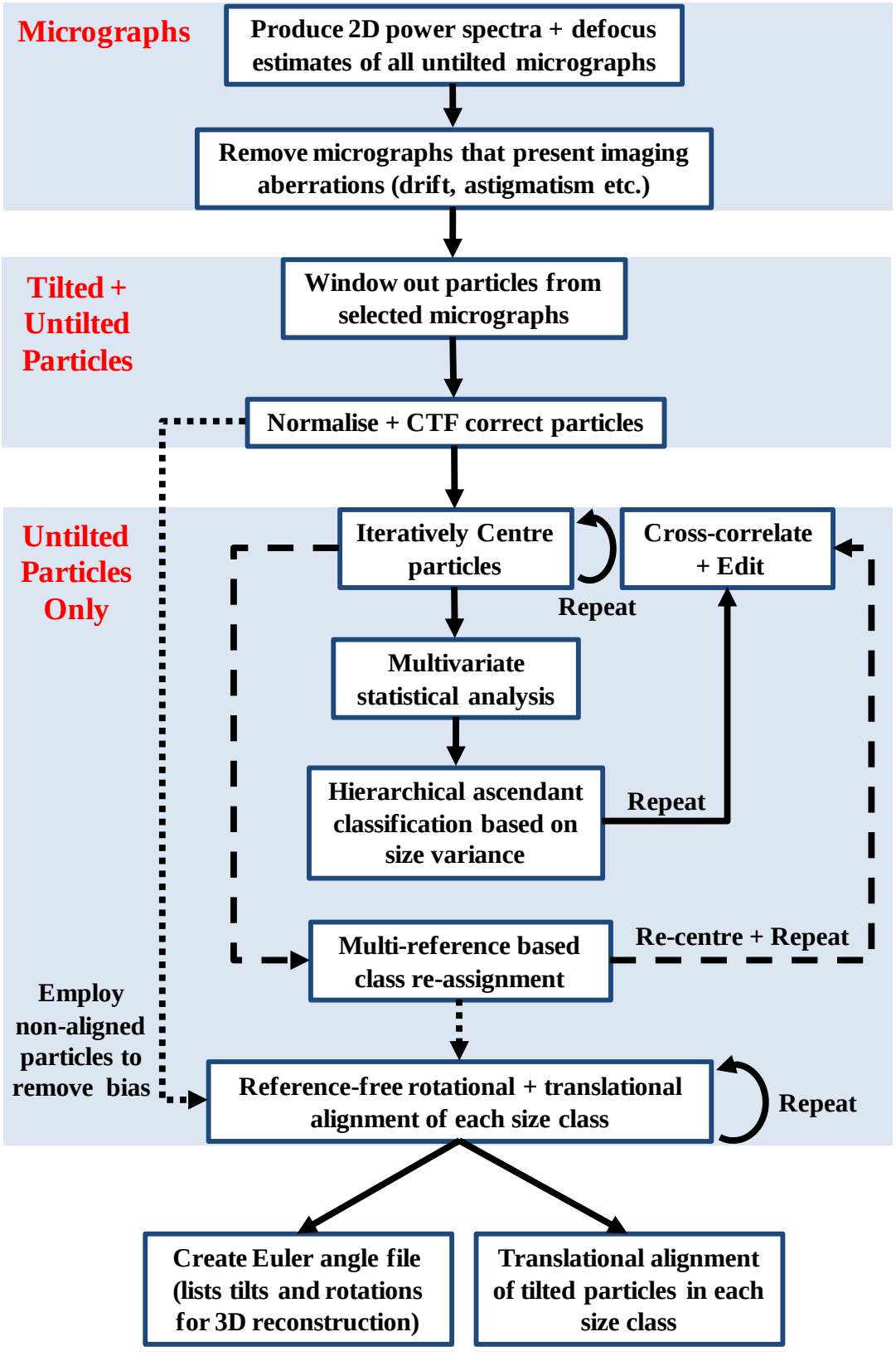
Contrast transfer function (CTF) estimation of tilted micrograph was performed using CTFTILT, and SPIDER based estimation of untitled micrograph CTFs were confirmed using CTFFIND3 (Mindell and Grigorieff, 2003).

All image processing procedures were performed on a OpenSUSE 11.1-2.6.27.7-9 Linux operating system. The programs were later migrated across to a Linux cluster to provide more hard drive space and reduce processing time by increasing the computing power.

This protocol presents a framework for the production of 3D density maps from single-particle EM data where the sample presents a preferred on-grid orientation and/or a size variation. The precise steps presented here are directly applicable to the data presented in chapter 4. The assessment of size variation presented here is based on an multivariate statistical analysis (MSA) method described previously (White *et al.*, 2004). This method was employed in the study of *E. coli* TatA complexes (Gohlke *et al.*, 2005) and proved successful for analysing and separating different sized particles.

### 2.10.2. Schematic overview of single-particle analysis

Schematic flow-diagrams outlining the main steps of the single-particle analysis of Tat complexes are shown below in Figure 3.2.1. The first of these details the 2D analysis of Tat complexes and the iterative procedure of size classification, whilst the second displays the steps involved in 3D reconstruction and refinement of the model. The steps are explained in further detail within the protocol.


Micrographs
Produce 2D power spectra + defocus estimates of all untilted micrographs
Remove micrographs that present imaging aberrations (drift, astigmatism etc.)
Tilted + Untilted Particles
Window out particles from selected micrographs
Normalise + CTF correct particles
Untilted Particles Only
Iteratively Centre particles
Cross-correlate + Edit
Repeat
Multivariate statistical analysis
Hierarchical ascendant classification based on size variance
Repeat
Multi-reference based class re-assignment
Re-centre + Repeat
Employ non-aligned particles to remove bias
Reference-free rotational + translational alignment of each size class
Repeat
Create Euler angle file (lists tilts and rotations for 3D reconstruction)
Translational alignment of tilted particles in each size class

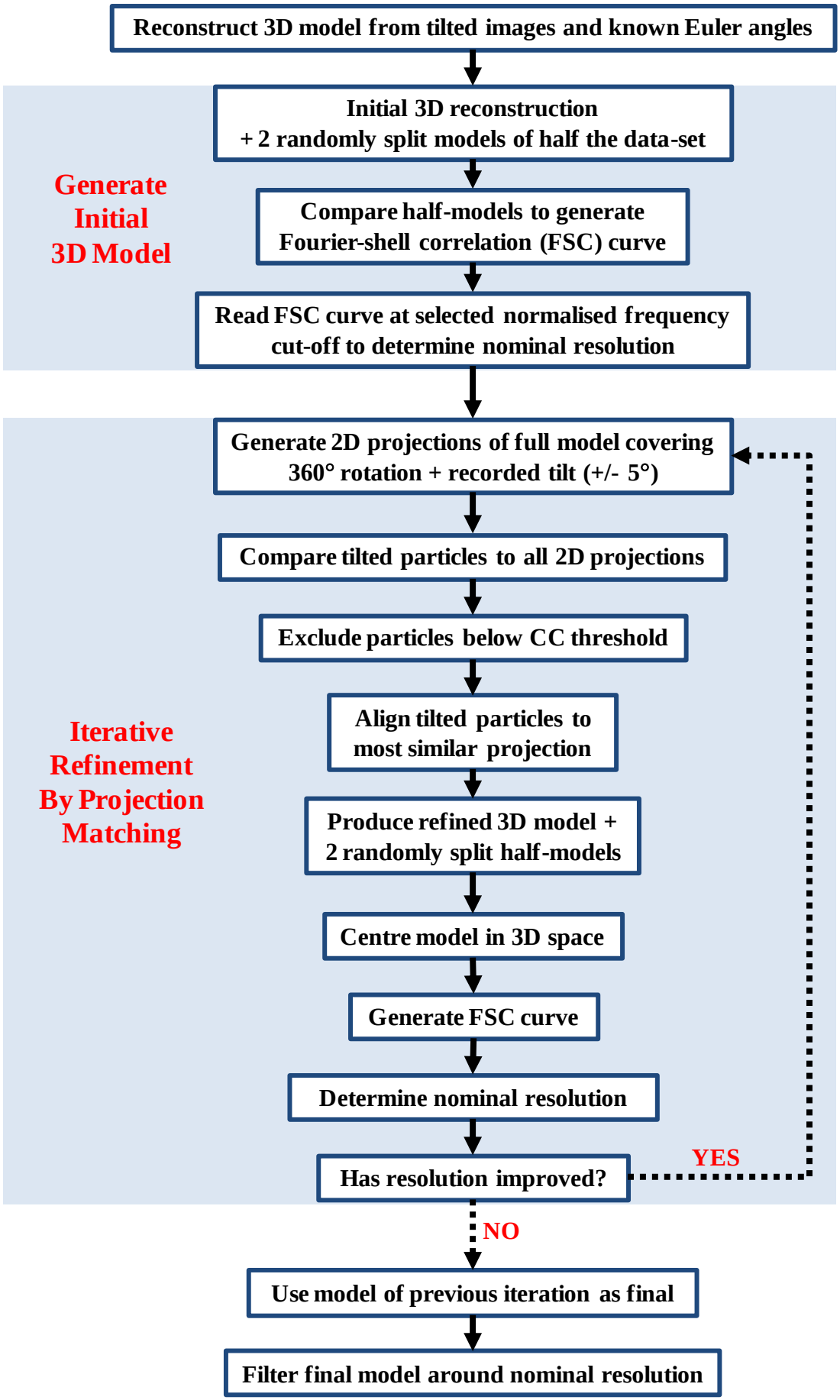

Reconstruct 3D model from tilted images and known Euler angles
Generate Initial 3D Model
Initial 3D reconstruction + 2 randomly split models of half the data-set
Compare half-models to generate Fourier-shell correlation (FSC) curve
Read FSC curve at selected normalised frequency cut-off to determine nominal resolution
Iterative Refinement By Projection Matching
Generate 2D projections of full model covering 360° rotation + recorded tilt (+/- 5°)
Compare tilted particles to all 2D projections
Exclude particles below CC threshold
Align tilted particles to most similar projection
Produce refined 3D model + 2 randomly split half-models
Centre model in 3D space
Generate FSC curve
Determine nominal resolution
Has resolution improved?
YES
NO
Use model of previous iteration as final
Filter final model around nominal resolution

**Figure 2.10.1 Schematic overview of 2D single-particle analysis of Tat complexes using a size variance separation method**
Full arrows represent the flow of information between stages. Dashed arrows represent a stage of the processing to be entered once sufficient iterations have been completed. Dotted arrows represent a stage of the procedure to be entered once certain thresholds have been met. CTF = Contrast transfer function, FSC = Fourier shell correlation, CC = Cross-correlation.

### 2.10.3. Protocol for size-based classification and 3D reconstruction using a random conical tilt (RCT) method.

All program names, directories and files formats or file extensions are indicated in bold text. The directory tree referenced is provided on the University of Warwick Department of Life Sciences Linux cluster designated **Coconut.**

*** =** The number of digits to use in the file name.

**I.** Run SPIRE **proj01** from the command line in **data** directory.

**II.** Open the dialog called **tilt3D** with all currently required batch files.

**III.** **<u>Micrograph selection and defocus determination</u>**

III.1. Select all the micrographs to be used and their tilted partners. These must all be at the same magnification. Treat the tilted and untilted micrographs as separate series.

III.2. Copy the selected micrographs (in **TIF** format) to a new **micrographs/TIF** directory and rename files as **Micro_**.spi** and **T_Micro_**.spi** (for the tilted series).

III.3. Run **copymrc.com** from the command line (runs **tif2mrc**) to copy **TIF** files to old **MRC** format (**CCP4**) and rename all output files as **.ccp4** files. Move the resulting **CCP4** files to a new **micrographs/CCP4** directory.

III.4. Run **T3D_tif_to_spi.cor** to convert **TIF** files to SPIDER files labelled **Micro_**.spi** and **T_Micro_**.spi**. Move the resulting **.spi** files to a new **micrographs/SPI** directory.

*III.5. Untilted micrographs*

III.5.1. Copy **ctffindall_Dan.com** to the **CCP4** directory and set the parameters, using the corrected "true" magnification (1.4313 times the recorded magnification). Run this from the command line. The output log files will contain defocus estimates +/- astigmatism.

III.5.2. Make sure the parameter file **param.spi** in **SPIRE** is correct, with magnification, pixel size, spherical aberration, amplitude contrast ratio and lambda fields filled (see *appendix 1* for these parameters).

III.5.3. Enter the number of tilt pairs into the **File numbers** box at the bottom of the **SPIRE** window (e.g. 1-12), hit enter to confirm.

III.5.4. Run **T3D_power.spi** on **Micro_**.spi** files (**SPIDER** format) to produce averaged power spectra for each micrograph. Eliminate any micrographs that show drift, astigmatism or low resolution and renumber the remaining micrographs and spectra consecutively.

III.5.5. Run **T3D_defocus.spi** to generate defocus values for the micrographs. The **defocusdoc.spi** output file contains defocus and astigmatism estimates for each micrograph in the series.

III.5.6. Compare the **ctffind** log files and **defocusdoc.spi** file (see *appendix 2* for converting **ctffind3.exe** results to **SPIDER** format). The defocus and astigmatism values should be similar. If not, try changing the defocus search

range in **ctffind.** If these still differ use the **ctffind** results and edit the **defocusdoc** accordingly, save as **final_defocusdoc**.

*III.6. Tilted micrographs*

III.6.1. Copy **ctftiltall_Dan.com** to the **CCP4** directory and set the parameters as for **ctffindall_Dan.com** and run this from the command line. The output **CTFTILT** log files will contain defocus, tilt angle and astigmatism estimates for the micrographs along with an equation for working out the defocus of each tilted particle based on its image coordinates. The final value for **TANGLE** (tilt angle) should be approximately the same as the recorded tilt from the microscope. If this is not the case try changing the parameters of **ctftiltall_Dan.com**.

III.6.2. Open **T3D_TiltDefocus.fed** and enter the required values which can be found at the end of the **CTFTILT** log. This will calculate the defocus values for each particle and print them into a document file. Repeat this for each micrograph.

*III.7. All micrographs*

III.7.1. After selecting the final micrographs to use, copy the **SPIDER** format images for each tilt pair into subfolders labelled **Pair_****. Each **Pair_**** folder should now contain an untilted micrograph labelled **Micro_**.spi** and a tilted micrograph labelled **T_Micro_**.spi**

(e.g **Pair_05** will contain **Micro_05.spi** and **T_Micro_05.spi**)

## IV. Particle picking and correction

IV.1. Run **Jweb** from the command line within the **Pair_01** subdirectory. Go to **File>Open>Image Series** and open both the untilted and tilted micrographs. Use **Options>Settings** to resize the micrographs so that both fit on the screen at once. On the **Image series viewer** open tab **Marker-Tilt pair** and click

**Show/Edit marker**. Pick several particles spread widely across the untilted micrograph and its tilted partner. To estimate the theta (tilt) angle of the micrograph click **Save marker file** then **Determine theta, Fit angles, Draw fitted locations** and finally **Save angles.** This process produces three output document files containing: Untilted image coordinates (**dcu_***.spi**), tilted image coordinates (**dct_***.spi**) and the angular determination file (**dcb_***.spi**).

IV.2. **Check** the **dcb_001.spi** document file, which can be found in the **Pair_01** directory, for the tilt angle under the **0124** key heading. This should be very close to both the angle recorded on the microscope and the angle given by **CTFTILT**. If not, close the **Image series viewer,** then reopen the tilt pair and re-pick 10-12 particles making sure the cursor is central on each pick. Then delete the dcb file and follow the steps to re-estimate the theta angle.

IV.3. Now that the relative orientations of the two micrographs have been confirmed, continue with the interactive pair-wise particle picking: after selecting the particle on the untilted micrograph the cursor should move automatically to the same particle on the tilted image. Avoid picking particles that are close to the edge of the micrograph or those that are bunched closely together. On average the particles should be separated by a particle's width from the next. Once all well separated particles have been picked, repeat the **theta angle** estimation as above to refine.

IV.4. Repeat these steps for each tilt pair with **Jweb** running within the corresponding **Pair_**** subdirectory. Change the **key number** for each **Pair_**** subdirectory you pick from, so that the particles are labelled as a continuous series.

IV.5. Run **T3D_mv_microdocs.spi** to move all the document files made by **Jweb** to a new document directory.

IV.6. Run **T3D_Tilt_box-out.fed** and then **T3D_Untilt_box-out.fed** to window-out the particles from the tilted and untilted micrographs respectively. This also generates particle list document files as particle-by-micrograph lists and as an entire particle series.

IV.7. Run **T3D_normalise.prj** on both the tilted and untilted series to normalise all particles and eliminate negative pixel values which will otherwise interfere with the correspondence analysis used to classify the particles later.

IV.8. Run **T3D_ctfcorr.cor** to correct the normalised untilted particles for the effects of the contrast transfer function (CTF). This is a mathematical function applied during image acquisition that artificially increases contrast over the micrograph allowing the user to distinguish particles from the background. However, this effect comes at the expense of data loss, as information over a certain frequency is subjected to increasingly rapid oscillations of alternating contrast. By computationally flipping the polarity of these frequency zones to produce a continuous positive transfer, the information content is restored to the whole resolution domain. This CTF correction method is called phase flipping.

IV.9. Run **T3D_tiltctfcorr.cor** to individually CTF correct each tilted particle using the defocus values from the **T3D_TiltDefocus.fed** document files.

IV.10. Run **T3D_filter.cor** on the untilted particles to apply a low-pass Butterworth filter to the CTF corrected particles. This will eliminate high frequency noise from the images with a gradual filter towards the higher resolution particle data, whilst preserving all the lower resolution structural information.

## V. Alignment, classification and generation of class averages

### *V.1. Tilted particles*

V.1.1. We want to centre the tilted particles without bias from any fringing artefacts introduced by the tilt. For this reason a custom mask can be applied to the tilted sum image as follows:

V.1.2. Run **T3D_centre.fed** on the tilted particle set to centre these particles without any rotation being applied. Run with no final mask.

V.1.3. Run **T3D_custom_mask.fed** with a tight initial mask radius and with **threshold ok?** set to **0.** Try a few threshold values until you get a good outline of the particle. Then set **threshold ok?** to **1** and run the batch again. Alter the **filter radius** and **final edge** to get a bigger or smaller final custom mask.

V.1.4. Then run **T3D_centre.fed** again with **use custom mask?** on. This time use the output particles from the initial centring run as the new input particles and add /**cm** (for custom mask) to the end of the output directory path.

*V.2. Untilted particles*

- **If a size range is present in the data-set then proceed from 5.2.1 to 5.2.3 and skip 5.2.4.**

- **If a size range is NOT present go directly to 5.2.4.**

V.2.1. Run **T3D_centre.fed** on the untilted particles to only align them translationally. Keep the initial mask fairly loose and don't mask the final particles. Check the centring by running **SPIDER** operation **CG** interactively from the command line. If the values for **x** and **y** are very close to 0.0 (e.g. 0.02) then proceed to 5.2.2. If not run **T3D_centre.fed** iteratively with a decreased mask radius to improve the centring.

V.2.2. Next remove any anomalous particles that do not centre well or have very low similarity to the average compared to the other particles. Use

**T3D_CC_TEST.spi** to cross-correlate the centred particles back to the centred average and use **T3D_plot_hist.spi** to view the results graphically. Record the spread and approximate peak for the cross-correlation coefficient (measure of similarity to the average). **T3D_CC_TEST.spi** removes the 10% of particles with the lowest cross-correlation value by default and saves the rest to a selection document file. Use **montagefromdoc** to evaluate the particles removed. If too many/few particles are removed then manually set the cut-off point and re-run **T3D_CC_TEST.spi** so only anomalous particles are removed.

V.2.3. Re-run **T3D_centre.fed** on the edited particle list, and then re-run **T3D_CC_TEST.spi** to check for an improvement in the cross-correlation. The spread of CCC values should have decreased and the approximate modal CCC value may be higher.

V.2.4. Run **T3D_APSR.cor** on the untilted particles. This will iteratively align the particles both rotationally and translationally using a reference-free based approach. The aligned particles will be masked and a final average image and variance image will be produced along with the document file listing the alignment angles.

V.2.5. Run **T3D_MSA.fed** on the aligned untilted particles to apply **MSA** (multivariate statistical analysis) by correspondence analysis. This generates a series of document files that define distances between the images based on their similarity in pixel densities. Also, a factor map is generated which displays the clustering of the images on a graph defined by the top 2 factors of pixel variance in the image set.

V.2.6. Run **T3D_Imp+Rec.fed** to generate a montage of importance (**MIMP**) and reconstituted (**MREC**) images based on the top 20 eigenfactors (variance factors in the image set). Each eigenimage is displayed as a pair of positive and negative images. The importance images show the pixel variance along the positive and negative portions of the factorial axis and the reconstituted images

show how an image would look that lies at the positive and negative extremes of each axis.

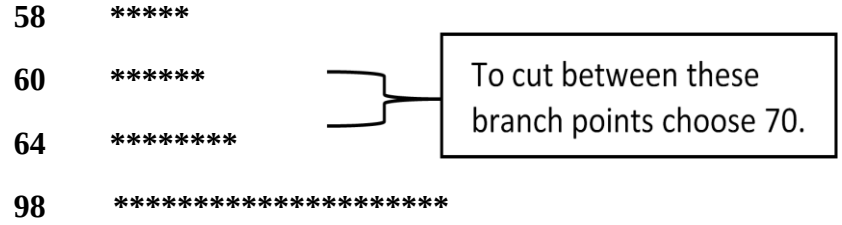
V.2.7. Analyse the **MIMP** and **MREC** images using **jweb** from the command line and identify which eigenimages display the main factors of variance in the dataset. From left-to-right the significance of the factors decreases, with the first eigenimage showing an overall sum image of the dataset. Make note of the most prominent variance factors showing different views or features.

V.2.8. If a size range is present in the particle set this will be clear from the importance images (a black and white halo of pixel variance around the rotationally averaged particle) and the reconstituted pair (a large and small particle). In this case note the number(s) of these eigenimages as you will initially have to classify the particles by size (see **hac.fed**), then rerun **MIMPs** and **MRECs** for each class to check if all size variance has been separated (i.e. these distinctive eigenimages are no longer present). If there is no longer any size variance in each class then you can run **T3D_APSR.cor** on each, and then proceed with the **MSA** to look for other significant variance factors, treating each size class separately throughout.

V.2.9. Run **T3D_hac.fed** to apply hierarchical ascendant classification to the aligned, untilted particles based on the selected eigenfactor(s). Unlike **k-means** classification, this is a self defining method of classification that isn't reliant on a user defined number of classes. First run on '**0'** setting to create the dendrogram document file and output the dendrogram readings to the **SPIDER** window interactively. View this output and note the maximum branch size (number followed by the most *s) and select a suitable cut-off value for the truncated dendrogram plot to define the number of classes generated. Look at the numbers above the max branch size and look for a large jump between the numbers (also shown graphically with the number of *s). Then choose an integer just above the lower of these. The lower the cut-off, the more classes generated. Then run the program again on '**1**' setting and input the MAX branch and chosen cut-off values. Also change the **HAC directory** to a

subdirectory labelled as the chosen cut-off (e.g. **CT_50**), this way when you try different cut-offs you don't have to rerun in '**0**' setting to reproduce the **dhac** doc file. This programme will generate class document files labelled **cldoc**.spi** and class average images.

**e.g.**




V.2.10.**T3D_hac.fed** will also generate an **sdvar** document file that displays the standard deviation of variance for each class. The lower this value, the greater the similarity of the particles in the class, and the better the classification. This should be compared to the number of particles in the class as well.

V.2.11.Run **T3D_rescalc.cor** to calculate the nominal resolution in angstroms for each class from the Fourier ring correlation (FRC). This provides an estimation of the class quality.

V.2.12.Try different **CT** (cut-off) values to alter the number of classes generated and improve the quality of the classification.

V.2.13.You can also run **montagefromdoc** from the command line to view the constituent particles in each class and edit out any that are broken or otherwise unwanted from each class. Save the new particle lists as **E_cldoc****.

V.2.14.Validation of the size separation can be performed by running **T3D_multi-ref.spi.** Here the centred class averages are used as cross-correlation references for the entire data-set, so each particle is allowed to find the most appropriate class. Then run **T3D_Inter-class_ptcl_tracker.spi** to produce particle lists showing the class re-assignments for each particle.

V.3. *For each class*

V.3.1. Rerun the **APSR** batch on the initial filtered particles using the class (or edited class) particle list document file (**cldoc**** or **E_cldoc****). This should improve the alignment as the smaller class particle set will be more homogeneous. Do not use the initially aligned particles (**apsr_m*****) as this will compound any alignment errors!

## VI. Generating a 3D model by back projection

VI.1. Run **T3D_Euler_doc.fed** to generate the Euler angles document file. This document file, along with the processed tilted particles, will be used to generate the 3D model. It is very important and must be correct to ensure an accurate model is produced. The 3 Euler angles (PSI, THETA and PHI) describe the rotations and tilts required to move each tilted particle into alignment and generate a 3D structure. Phi is the in-plane rotation used to align the untilted particles, theta is the tilt angle relating the tilted and untilted particles and psi is default set to zero.

VI.2. Run **T3D_Intial_3D.fed** on the unmasked tilted particles to generate an initial 3D reconstruction using the faster but less accurate back projection method **BP 32F** (set parameter x15=1). This method will produce three 3D structures. The first will be a full model using all particles in the data-set, and the other 2 will each use half of the particles divided randomly, labelled ***_odd** and ***_eve.**

VI.3. The 3D structures can be viewed using **Chimera** (Pettersen *et al.*, 2004) or other 3D viewing software.

VI.4. Run **T3D_FSC.spi** to calculate the similarity between the 2 half models based on the Fourier Shell Correlation (FSC) (Harauz and van Heel, 1986; Radermacher, M. *et al.*, 1987; Radermacher, M. *et al.*, 1987). This compares the half-set reconstructions in Fourier space resulting in a measure of correlation between 1 (complete similarity) and 0 (no similarity) across the normalised

frequency range. When viewed as a 1D plot this graph should show high similarity at low frequency dropping to 0 as the frequency increases. Then run **T3D_Res.spi** to calculate the resolution of the initial model based on the 0.5 FSC cut-off (Bottcher *et al.*, 1997).

VI.5. Run **T3D_Refine_3D.fed** to perform an iterative refinement of the 3D structure by projection matching. This is a long procedure with a range of variables for optimisation. First 2D projections of the initial model are generated covering a full rotation of the structure tilted to the recorded Theta angle +/- 5° (To account for any localised variations in tilt across the grid). The tilted particles are then cross-correlated to all the projections to find the most similar match. Any particles with a CC below a set threshold can be excluded as anomalous. The selected particles are then aligned to the most similar projection and a refined 3D model and half-set models are produced. The refined model is then centred in 3D space to avoid any cumulative centring errors over the iterations. The resolution is then estimated using the same 0.5 FSC method and this is compared to the previous resolution estimate. If the resolution has improved the structure is used as the starting model for the next round. The procedure is iterated until the resolution no-longer improves or the iteration limit is reached. Check the FSC and if the curve increases at high frequencies after an initial decrease try increasing the value for Lambda (used to control speed of convergence in **BP RP and BP CG** reconstruction methods). However, if this is set too high the back projection will terminate early before the requested number of iterations has been performed and the 3D structure has converged. This can be seen on the SPIRE readout. In this case reduce the value of lambda. Also try changing the radius for the projections, translational search range and the first and last rings for alignment to improve the refinement. Finally run **T3D_filt.spi** to filter the final structure using a Butterworth low-pass filter around the nominal resolution, as estimated from the FSC 0.5 cut-off.

# *Chapter 3*

# *Structural analysis of TatAd complexes*

## 3.1. Introduction

Under steady state conditions the three essential components of the *E. coli* Tat pathway are seen to form two primary types of integral membrane complex: a TatABC substrate-binding complex of ~ 370 kDa and TatA complexes ranging in size from 50 kDa to over 500 kDa (Bolhuis *et al.*, 2001; Oates *et al.*, 2005; Oates *et al.*, 2003). The TatABC complex and TatA complex or complexes are thought to transiently coalesce to form the active translocon, and TatA has been suggested to form pores of varying size that could accommodate different substrates. Low resolution EM structures of purified TatA suggest that this protein assembles into ring-shaped particles of 9-13 nm diameter, a small subset of which have an internal cavity that could be large enough to accommodate the larger Tat substrates (Gohlke *et al.*, 2005).

On the other hand, data presented later in this manuscript (chapter 4) have shown that TatE, which appears to fully substitute for TatA, forms complexes that are much smaller and more homogeneous. These complexes appear as rings of 6-8 nm (Baglieri *et al.*, 2012), which are too small to accommodate large Tat substrates in a folded state. The precise nature and function of TatA-type complexes is thus currently unclear.

In contrast to Gram-negative bacteria, almost all Gram-positive bacteria possess a 'minimalist' Tat system which lacks a TatB component. Interestingly some Gram-positive organisms contain multiple Tat systems which possess differing substrate specificities (Jongbloed *et al.*, 2006). *Bacillus subtilis* is one such bacterium, containing two minimal AC-type complexes, termed AdCd and AyCy, along with a third TatA component, TatAc, the function of which has so far remained unclear. The *tatAd* and *tatCd* genes are expressed under phosphate-limiting conditions and are located together in an operon downstream of the *phoD* gene; this encodes a protein with both phosphodiesterase and alkaline phosphatase activity, and is the only known substrate of the TatAdCd translocase (Jongbloed *et al.*, 2004; Jongbloed *et al.*, 2000; Pop *et al.*, 2002). The absence of a TatB component is compensated for by the bifunctional role of the TatAd protein, which has been shown to complement both *E. coli tatA/E* and *tatB* null mutants (Barnett *et al.*, 2008).

As with the *E. coli* system the *B. subtilis* Tat components have been shown to form two types of complexes; a TatC-containing complex that we assume to be functionally analogous to the TatABC complex, and a separate TatA complex. However, the *B. subtilis* complexes are significantly smaller and more homogeneous than the *E. coli* versions. The TatAdCd complex runs at ~ 230 kDa during Blue-native PAGE and the TatAd complex has been estimated to be ~ 160 kDa by gel filtration (Barnett *et al.*, 2008), while both TatAyCy and TatAy have been reported to form ~200 kDa complexes as judged by gel filtration (Barnett *et al.*, 2009; van der Ploeg *et al.*, 2011) These earlier studies also showed that the TatAdCd system is able to export the large, cofactor-containing Trimethylamine N-oxide (TMAO) reductase (TorA) substrate (~ 90 kDa) when expressed in an *E. coli tat* null mutant. These data therefore suggest that multiple-sized TatA complexes are not essential for the effective transport of Tat substrates. Consistent with these findings, an alternative model for Tat translocation suggests that a localised concentration of TatA components acts to destabilise the lipid bilayer to facilitate transport (Brüser and Sanders, 2003).

A further area in which the *B. subtilis* Tat systems appear to differ from that of *E. coli* is in the presence of a cytoplasmic species of TatA. Cytoplasmic TatAd has been reported to bind the PhoD substrate (Pop *et al.*, 2003) and form homo-multimeric complexes varying widely in both size and shape (Westermann *et al.*, 2006). Similarly large complexes or aggregates (~5 MDa) were identified for a cytoplasmic TatAy population (Barnett *et al.*, 2009). It has been suggested that cytoplasmic TatAd acts as a Tat substrate chaperone shuttling to the membrane localised TatCd component (Schreiber *et al.*, 2006); however the biological significance of the cytoplasmic population has remained controversial.

Despite the clear functional overlap between the Tat systems of *E. coli* and *B. subtilis*, the available evidence suggests major differences in complex organisation and potentially in the translocation mechanism itself.

In this chapter the first structural investigation of membrane-localised *B. subtilis* TatAd complexes is presented.

Initially, the results of optimisation experiments to determine on-grid imaging conditions are presented. These studies demonstrate the effects of detergent concentration on the appearance of the observed particle population.

The results of Nanogold binding experiments are presented to validate the identification of the particle population and provide direct evidence for a modular arrangement of TatAd subunits within the observed particles.

Finally the results of a single-particle EM analysis of TatAd complexes are presented This study includes an investigation into particle size variance and the production of the first 3D density maps of TatAd complexes. These data show that although TatAd complexes do form ring-shaped structures, these appear to be smaller and more homogeneous than *E. coli* TatA complexes, with no indication of a pore large enough to translocate the larger Tat substrates in a folded state.

## 3.2. Purification of TatAd-*his* complexes in DDM

### 3.2.1. Membrane isolation and affinity purification of TatAd-*his*

TatAd-*his* was over-expressed in *E. coli ΔtatABCDE* cells from the pBAdh plasmid as described previously (Barnett *et al.*, 2008). 1.5 L of cells was cultured to mid-exponential growth phase before induction with 500 µM arabinose. The cells were fractionated as before, the membranes isolated and solubilised in 1% dodecyl maltoside (DDM). This detergent was selected due to its small micelle size and low critical micelle concentration (CMC) making it suitable for analysis by electron microscopy. The solubilised membranes were applied to 10 ml pre-equilibrated Talon™ slurry and left rotating overnight at 4°C to allow for optimal binding to the entirety of the resin. The column was washed with 3 x 20 ml of equilibration buffer (5 mM imidazole) before elution across 20 x 1.4 ml fractions using buffer supplemented with 150 mM imidazole. Figure 3.2.1 shows a silver-stained gel of the Talon™ column fractions and the corresponding immunoblot against the His-tag. From this figure the majority of TatAd-*his* can be seen to elute over fractions 7-20 with some lost in the initial flow-through and wash fraction. Some contaminating bands can also be seen.

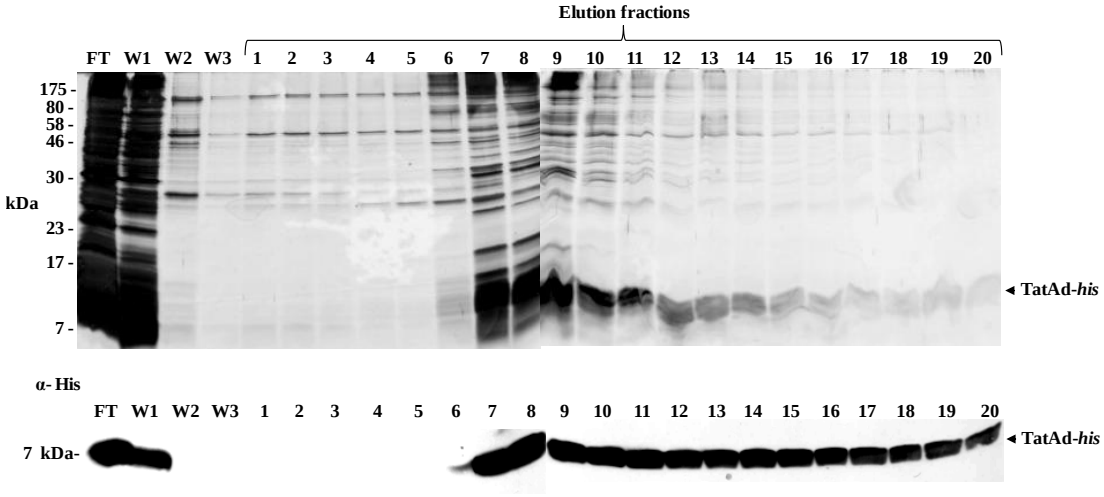



**Figure 3.2.1. Affinity chromatography of TatAd-*his***
*E. coli* membranes were isolated from *ΔtatABCDE* cells expressing TatAd-*his* from the pBAdh plasmid and solubilised in DDM. Membranes were applied to a Talon™ column as described above. The column flow-through (FT), wash fractions (W1-3) and elution fractions (elution 1-20) were analysed by silver-stain (top panel) and immunoblotted with antibodies against the *His*-tag on TatAd. Positions of TatAd are indicated on the right of the figure and molecular weight markers are shown on the left.

All Tat-containing elution fractions were pooled and concentrated using a 30 kDa molecular weight cut-off (MWCO) concentrator. As shown in Figure 3.2.2, no protein was detected in the flow-through during this procedure, suggesting that multimeric complexes are present.

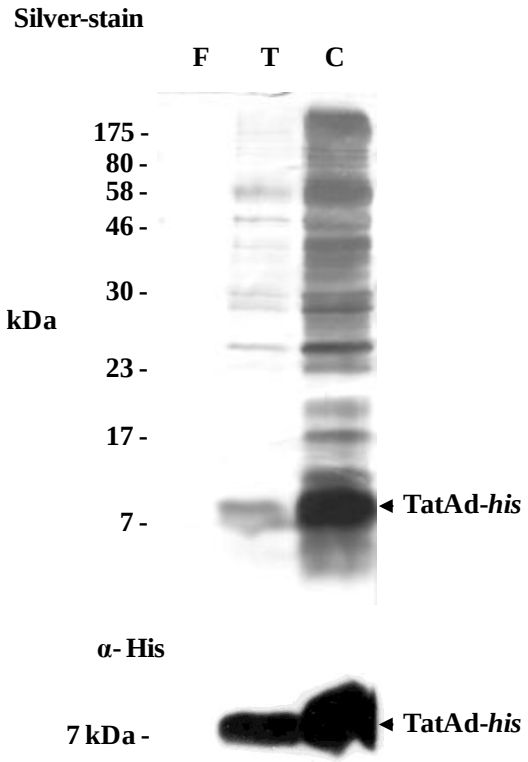
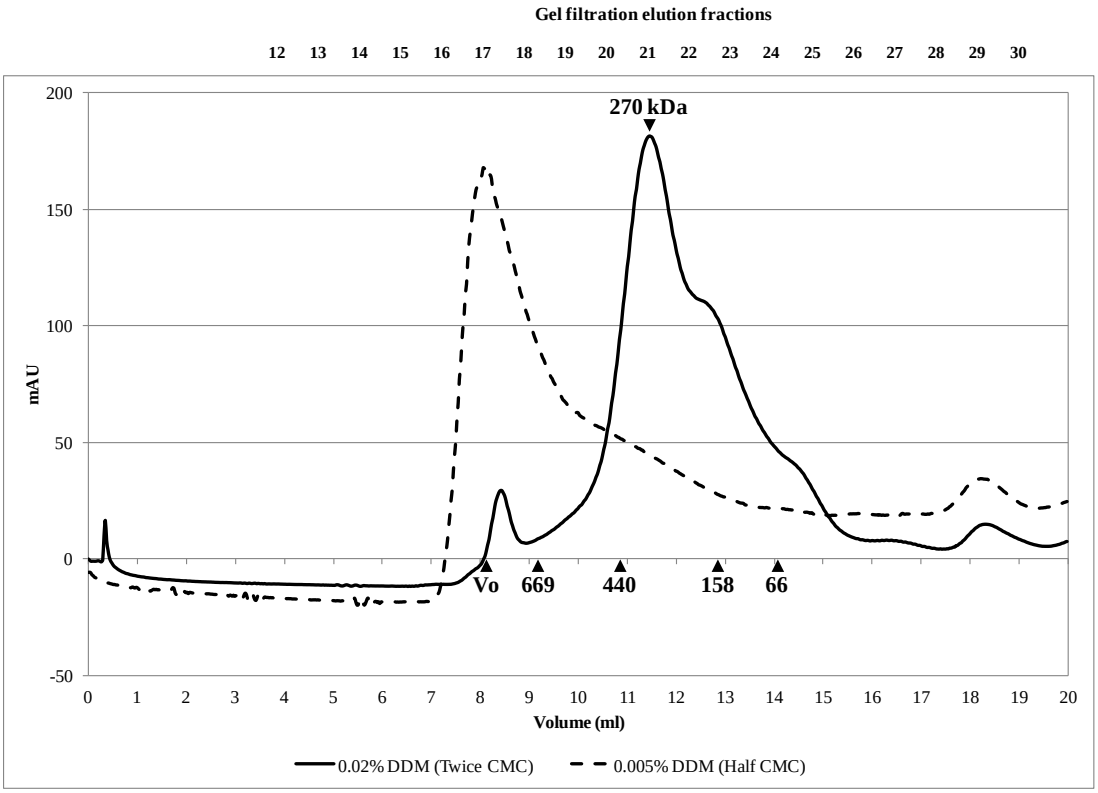



**Figure 3.2.2. Concentration of TatAd-*his* complexes**
TatAd containing Talon™ column elution fractions 7-20 were pooled (~ 20ml) and concentrated 10x using Vivaspin 20 MWCO 30 kDa concentrators. The filtrate (F), total pooled protein (T) and concentrate (C) were analysed by silver-stain and immunblotted with antibodies against the *his*-tag on TatAd. Positions of TatAd are indicated on the right of the figure and molecular weight markers are shown on the left.

### 3.2.2. TatAd-*his* gel filtration profile under differing detergent concentrations

A 240 µl sample of the concentrate was loaded onto a calibrated Superdex™ 200 10/300 GL gel filtration column, pre-equilibrated with GF buffer (20 mM Trizma, 150 mM NaCl) supplemented with 0.02% DDM. This concentration of DDM was used previously to purify TatAd-*his* complexes, and converts to ~ 2x the detergent CMC (CMC of DDM in water = 0.009% w/v, Thermo Scientific). After elution of the sample over a full column volume using 0.5 ml fractions, the column was re-equilibrated with GF buffer + 0.005% DDM (~ ½ the detergent CMC). A second aliquot of TatAd-*his* concentrate was then applied to the column and eluted using this buffer. Figure 3.2.3 shows the 220 nm UV trace across both of these runs along with

the position of soluble protein standards of known weight. When DDM is present at twice the CMC, TatAd-*his* elutes across the resolving range of the column with a well defined peak corresponding to ~ 270 kDa, relating to elution fraction 21. From the immunoblot, TatAd-*his* can be clearly seen across fractions 20-26. However, when DDM is present at half the CMC, a major change in elution profile is seen. The vast majority of TatAd-*his* is now present in the void volume (Vo), indicating that the sample has aggregated and can no longer enter the matrix of the Superdex™. This observation is confirmed by the Western blot which shows a peak of TatAd-*his* in fractions 14+15 (corresponding to the void) trailing across the separation range of the column. These data indicate that 0.02% DDM is sufficient to maintain TatAd-*his* complexes in a soluble state.



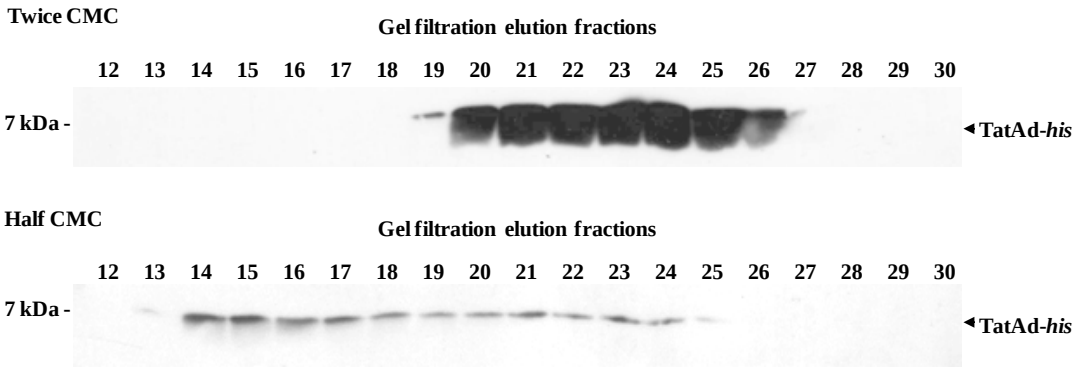




**Figure 3.2.3. Gel filtration of TatAd-*his* under differing DDM concentrations**
A 240 µl sample of TatAd-*his* concentrate was applied to an equilibrated Superdex™ 200 10/300 GL gel filtration column as described above. The top panel shows the 200 nm UV trace for sample runs using twice CMC DDM (full line) and half CMC DDM (dashed line) supplemented GF buffer. The elution points for protein standards of known weight are shown below along with the column void volume (Vo). The lower panel displays immunoblots against the *his*-tag on TatAd. The elution fractions shown correspond to a column volume of 8-15 ml. Positions of TatAd are indicated on the right of the figure and molecular weight markers are shown on the left.

Gel filtration fraction 21 was selected for visual analysis by electron microscopy and the purity of the sample was assessed by silver-stained SDS-PAGE. As shown in Figure 3.2.4 TatAd-*his* represents the dominant band with little contaminating protein. The protein concentration of this fraction was estimated at 24 µg/ml using Bradford's reagent (see methods).

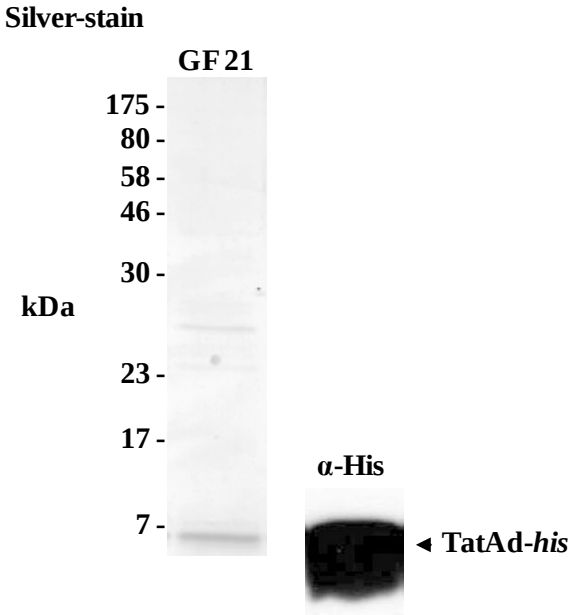



**Figure 3.2.4. Silver-stain SDS-PAGE of purified TatAd-*his* for EM analysis**
The purity of gel filtration elution fraction 21 was assessed by silver-stain. The corresponding immunoblot is also shown. The Position of TatAd is indicated on the right of the figure and molecular weight markers are shown on the left.

## 3.3 Optimisation of EM grid conditions for TatAd-*his* complex visualisation

In order to obtain meaningful images of TatAd-*his* complexes by EM, on-grid conditions were investigated so as to provide an optimum imaging environment. Various sample dilutions were trialled to provide an even distribution of well separated particles across the grid. Staining duration and concentration were also investigated to provide suitable contrast of the particles against the carbon background, whilst maintaining resolution of internal density variation across the particles. Detergent-solubilised proteins are present in solution as a mixed population of particles. These consist of protein-detergent-complexes (PDCs) and detergent micelles. Therefore washing steps were introduced to remove excess detergent from the grid before imaging. Using standard EM techniques for single-particle analysis it is not possible to specifically distinguish between these types of particle or to determine the detergent-to-protein ratio of a PDC. As such, conditions were investigated so as to reduce the detergent concentration to the minimal amount required to solubilise the membrane protein. Under these conditions the PDCs should consist of mostly protein and the population of free micelles should be somewhat reduced.

### 3.3.1. Single particle EM of TatAd-*his* undiluted

Initially an undiluted sample of fraction 21 was used. After sample addition the grid was washed twice with detergent-free GF buffer before staining twice with 2% uranyl acetate. As shown in figure 3.3.1 the grid is overcrowded with a mixture of small round particles (~ 8 nm diameter), the majority of which display reduced central density indicating a cavity or channel. Some clouding and aggregation can also be seen due to the high sample concentration. Sample dilution was therefore required to provide a suitable distribution of complexes for single-particle analysis.

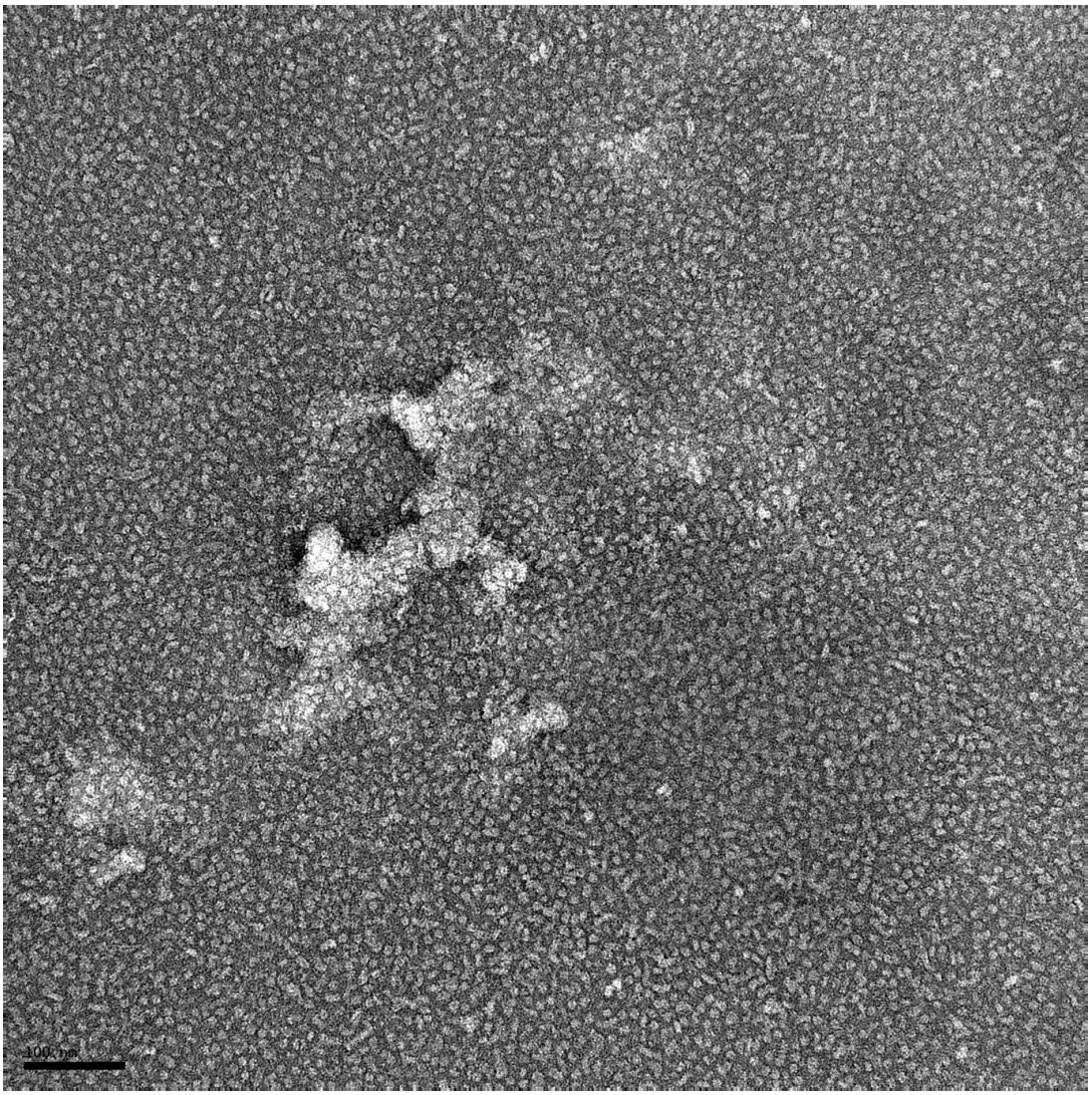



**Figure 3.3.1. Negative-stain micrograph of TatAd-*his* undiluted**
TatAd-*his* complexes are present as a continuous lawn of small ringed shaped particles. Some clouding and aggregation of the sample can be seen. The micrograph was taken at 57252x magnification under ~1.5 µm defocus. The grid was stained with 2% uranyl acetate. Scale bar = 100nm

### 3.3.2. Single-particle EM of TatAd-*his* 1:4 dilution in detergent-free GF buffer

An aliquot of fraction 21 was diluted 1:4 in buffer lacking DDM prior to gridding as before. Under these conditions a major shift in particle morphology was observed, as shown in figure 3.3.2. Although the particles now appear well dispersed, several distinct populations can now be seen that were absent in the undiluted sample. These consist of large amorphous structures of over 50 nm in diameter (Figure 3.3.2, arrows); short rod-shaped particles of up to 25 nm in length (Figure 3.3.2, dashed arrows); and small round particles present in a range of sizes from less than 8 nm to over 20 nm in diameter (Figure 3.3.2, arrow heads). As the detergent is now present at ~ ½ the CMC (0.005% DDM) none of these particles should represent detergent micelles. This conclusion is confirmed by Figure 3.3.3, which displays a negative control of only GF buffer + 0.005% DDM. In this case no particles were visible on the grid.

In light of the gel filtration data shown previously (Figure 3.2.1) it can be concluded that the majority of TatAd-*his* complexes shown in Figure 3.3.2 are no longer stable and have collapsed into a disordered array of aggregates. However, the presence of the smaller, round particles (Figure 3.3.2, arrow heads) suggests that localised concentrations of DDM are enough to solubilise a proportion of the TatAd-*his* complexes as PDCs with varying detergent contribution. Overall these diverse particle populations correlate well with the vast size distribution of TatAd-*his* seen via gel filtration for the ½ CMC DDM sample (Figure 3.2.3). This is estimated to be > 1500 kDa to < 50 kDa based on the soluble protein standards.

Interestingly this heterogeneous mix of particles closely resembles EM images of the contentious soluble population of TatAd complexes observed previously (Westermann *et al.*, 2006). In this case a similar size distribution and range of particle populations were recorded for over expressed TatAd that localised in the cytoplasm during fractionation. The similarities between these imaging data indicate that such a soluble population likely represents aggregation of excess Tat protein.

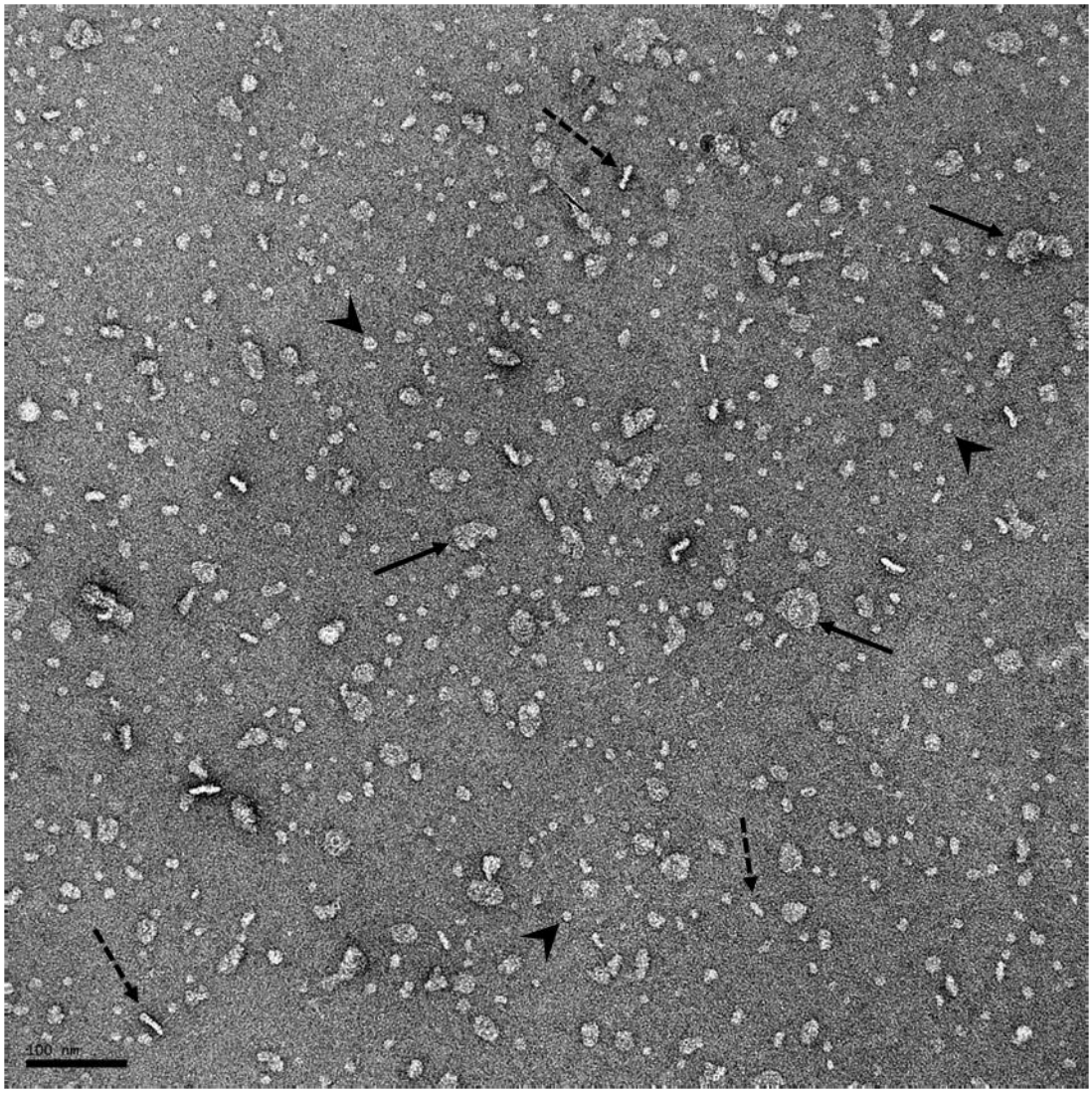



**Figure 3.3.2. Micrograph of TatAd-*his* 1:4 dilution in detergent-free GF buffer**
The micrograph was taken at 57252x magnification under ~1.5 µm defocus. The grid was stained with 2% uranyl acetate. Distinct particle populations are indicated as follows: amorphous aggregates (arrows), rod-shaped particles (dashed arrows) and small, round particles (arrow heads). Scale bar = 100nm

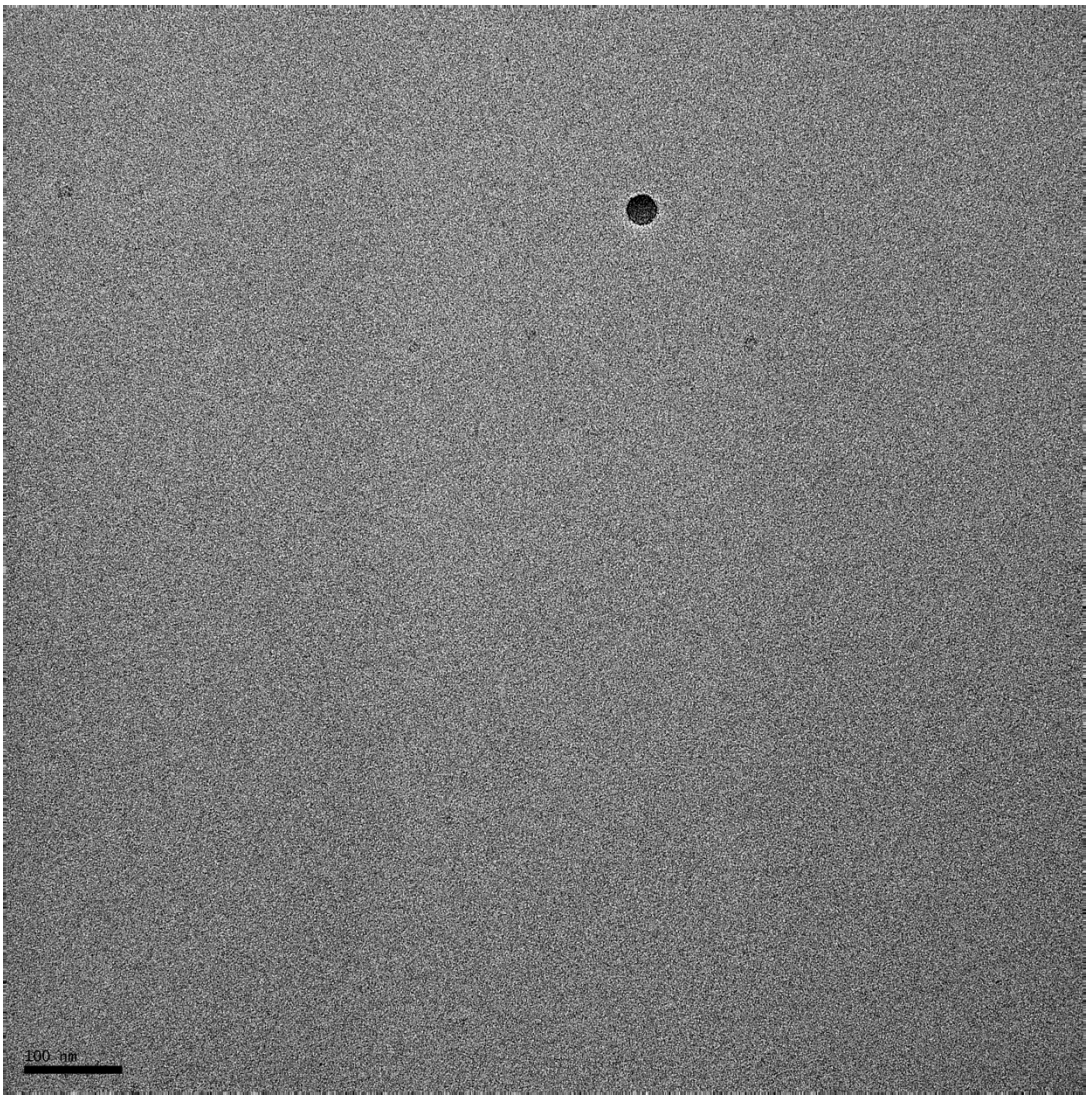



**Figure 3.3.3. Micrograph of half CMC DDM (0.005%) negative-control grid.**
The micrograph was taken at 57252x magnification under ~1.5 µm defocus. The grid was stained with 2% uranyl acetate. No particles were identified; the large black circle shown is a grid artefact. Scale bar = 100nm

### 3.3.3 Single-particle EM of TatAd-*his* 1:4 dilution in GF buffer + 0.02% DDM

When the same 1:4 dilution of TatAd-*his* is performed using the same GF buffer supplemented with detergent, the undiluted particle phenotype is restored, as shown in Figure 3.3.4. Some aggregates and broken particles can still be seen but the grid is once again dominated by small, round particles. This demonstrates that the aggregated particles seen in Figure 3.3.2 are attributable to the change in DDM concentration and not protein concentration. The distribution of particles shown here is not optimal for single particle analysis, so further dilutions were performed.

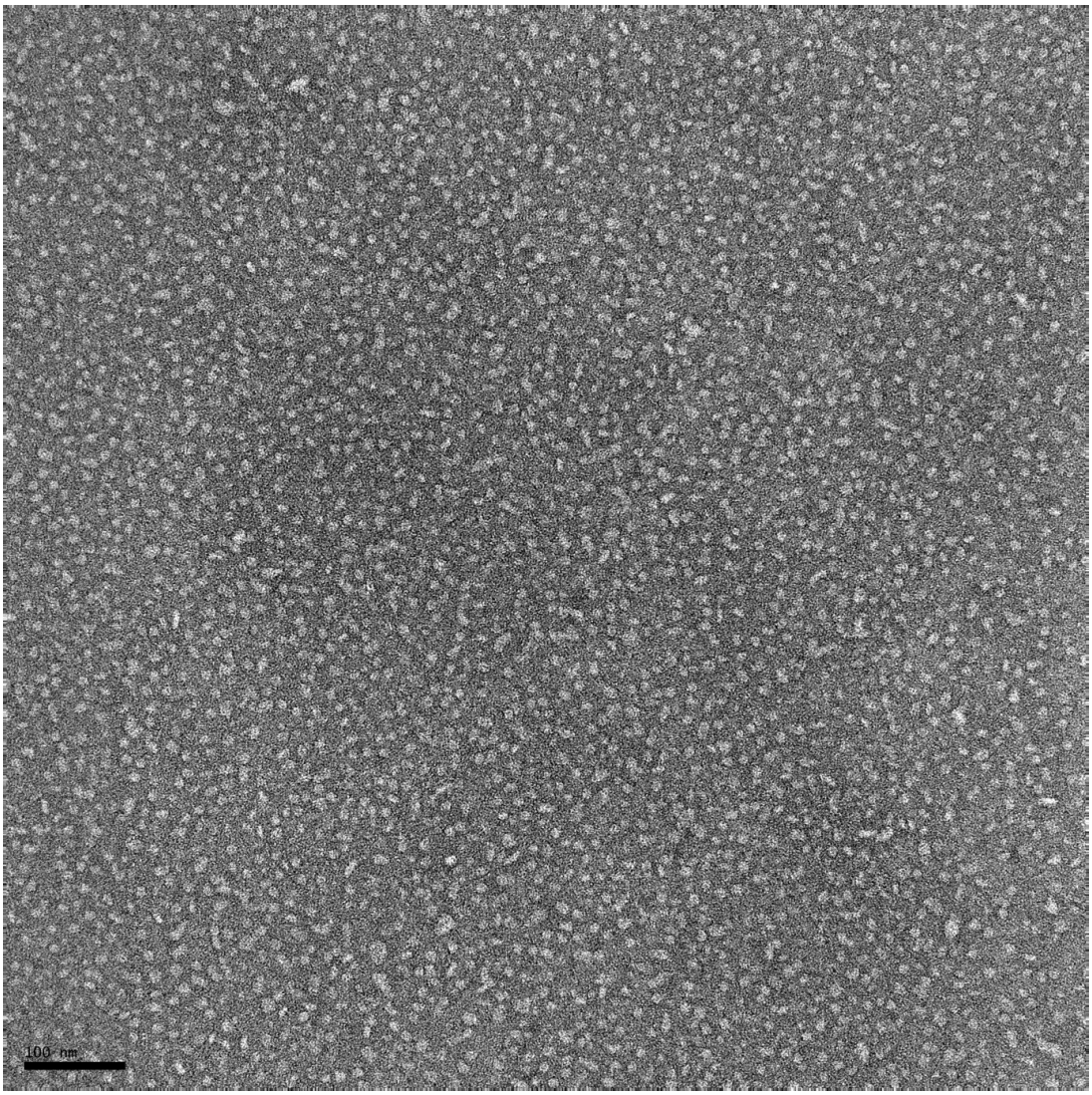



**Figure 3.3.4. Micrograph of TatAd-*his* 1:4 dilution in GF buffer + 0.02% DDM**
The micrograph was taken at 57252x magnification under ~1.5 µm defocus. The grid was stained with 2% uranyl acetate. Scale bar = 100nm

### 3.3.4. Single-particle EM of TatAd-*his* 1:8 dilution in GF buffer + 0.02% DDM

A final 1:8 dilution of TatAd-*his* using the GF buffer + 0.02% DDM resulted in the distribution of particles as shown in Figure 3.3.5; this corresponds to a protein concentration of ~ 3 µg/ml. In this case the particles appear well separated with a clear boundary of stain encircling each. The carbon background is not overly stained, providing a good level of contrast, leaving the protein complexes easily distinguishable. A clear central pool of stain can be seen in many of the particles. The imaging conditions and particle distribution observed for this dilution was deemed suitable for single particle analysis, as demonstrated by the montage of well defined, unprocessed particles shown in Figure 3.3.6.

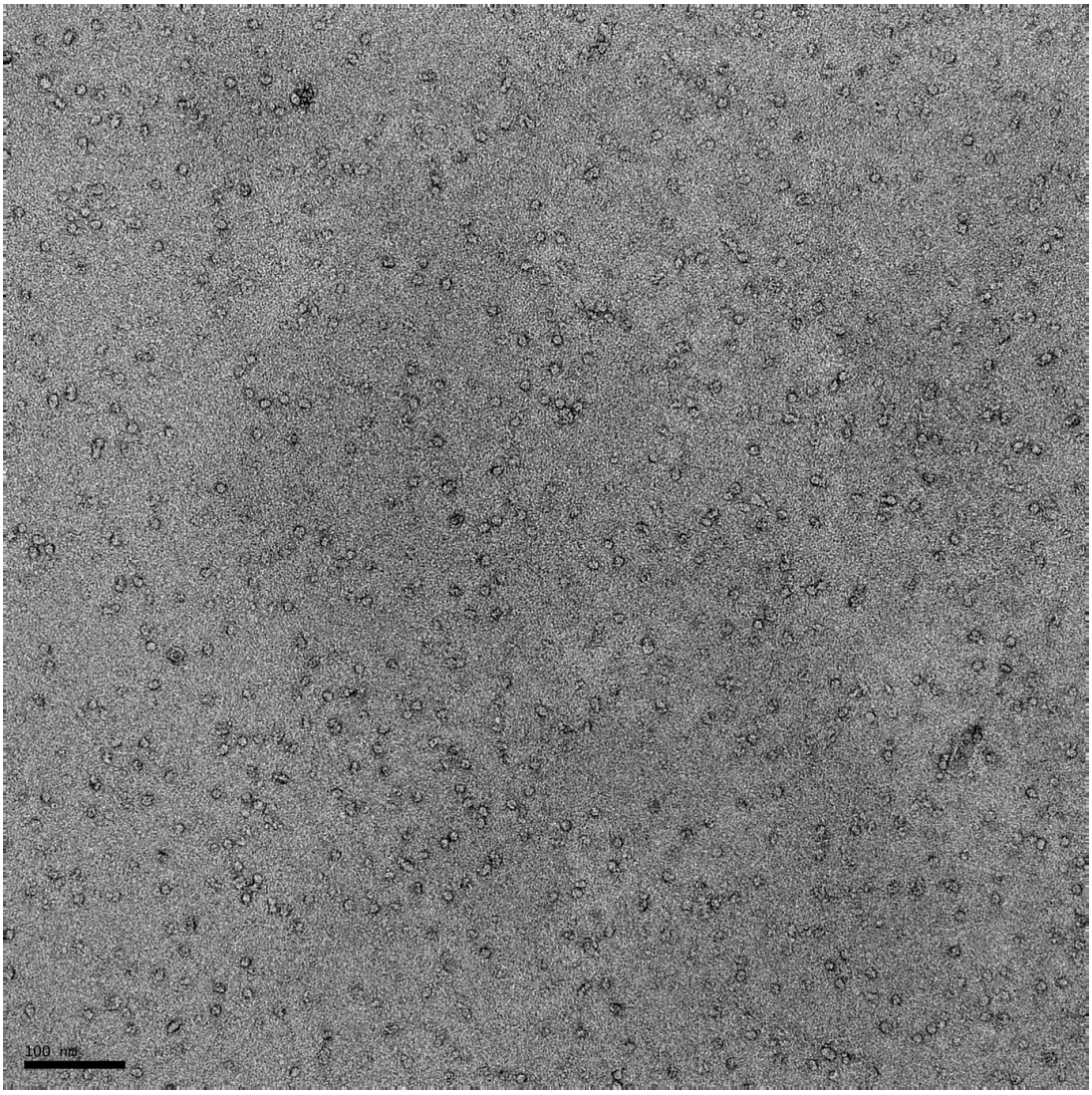



**Figure 3.3.5. Micrograph of TatAd-*his* 1:8 dilution in GF buffer + 0.02% DDM**
The micrograph was taken at 57252x magnification under ~1.5 µm defocus. The grid was stained with 2% uranyl acetate. Scale bar = 100nm

Manual size estimation of these windowed-out particles gives an average diameter of ~ 7.5 nm based on 2.62 Å/p resolution at 57252x magnification. A more accurate single particle assessment of these complexes is provided later in the chapter.

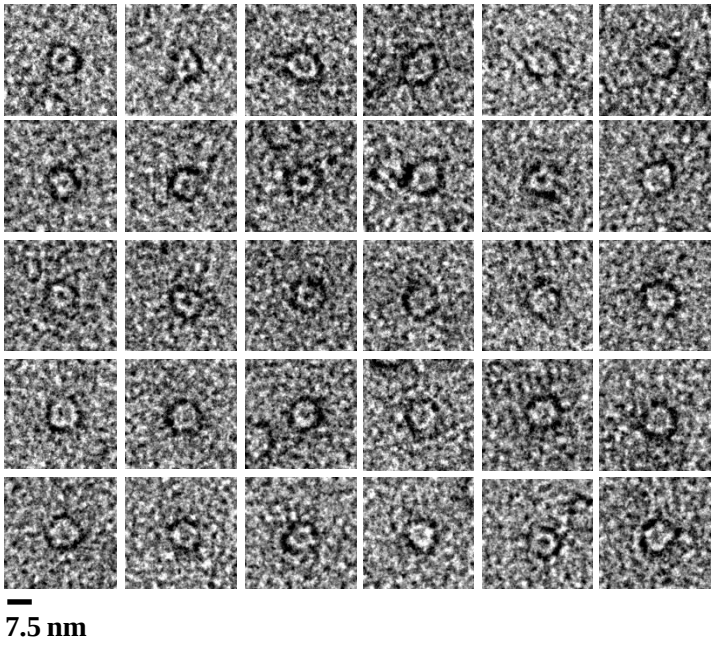



**Figure 3.3.6. Montage of TatAd-*his* complexes in DDM under negative-stain**
Example particles taken from the 1:8 dilution in GF buffer + 0.02% DDM. Particles were picked, using the EMAN suite program Boxer, into 128 x 128 pixel windows. Raw (unprocessed) images are shown. Particles are small and roughly circular measuring ~ 7.5 nm in diameter. A central stain pool can be seen in each particle indicating a cavity or channel.

## 3.4. Nanogold labelling of TatAd-*his*

Having concluded the optimal sample dilution for adequate particle dispersion on the grid, further imaging experiments could be conducted. There is an inherent bias imposed by the user when selecting particles for image processing; by either interactively picking them or editing a computer selection. In light of the observed effects of detergent concentration on particle morphology, further validation of the particles shown in Figure 3.3.6 would be desirable.

As such, specific molecular labelling was used to validate the identity of the particle population and to investigate the stoichiometry and conformation of subunits within

the complex. 1.8 nm Ni-NTA-Nanogold® (Nanoprobes) was employed for this purpose due to the high affinity for metal ions displayed by His-tagged proteins. This label consists of a 1.8 nm Nanogold particle with multiple nickel-nitrilotriacetic acid functionalities attached. Each $Ni^{2+}$ coordinates with one nitrilotriacetic acid and two histidines on the His-tag of TatAd-*his*. These particles are easy to detect by EM due to the very electron dense gold that appears as an intense black dot.

### 3.4.1. TatAd-*his* + 1:10 Ni-NTA-Nanogold®

For the initial labelling experiment, the 1:8 TatAd-*his* grid preparation method was used, but with the addition of a 15 minute on-grid incubation with 1:10 Nanogold (1 nmol/ml), diluted in GF buffer + 0.02% DDM, after the 1 minute TatAd sample incubation. From Figure 3.4.1 a high concentration of TatAd-*his* complexes can be seen along with a great number of Nanogold particles (intense black dots). The contrast appears very low however, and is greatly limiting particle resolution; the high Nanogold concentration may be affecting this. Further optimisation of imaging conditions was clearly required to sufficiently resolve the particles.

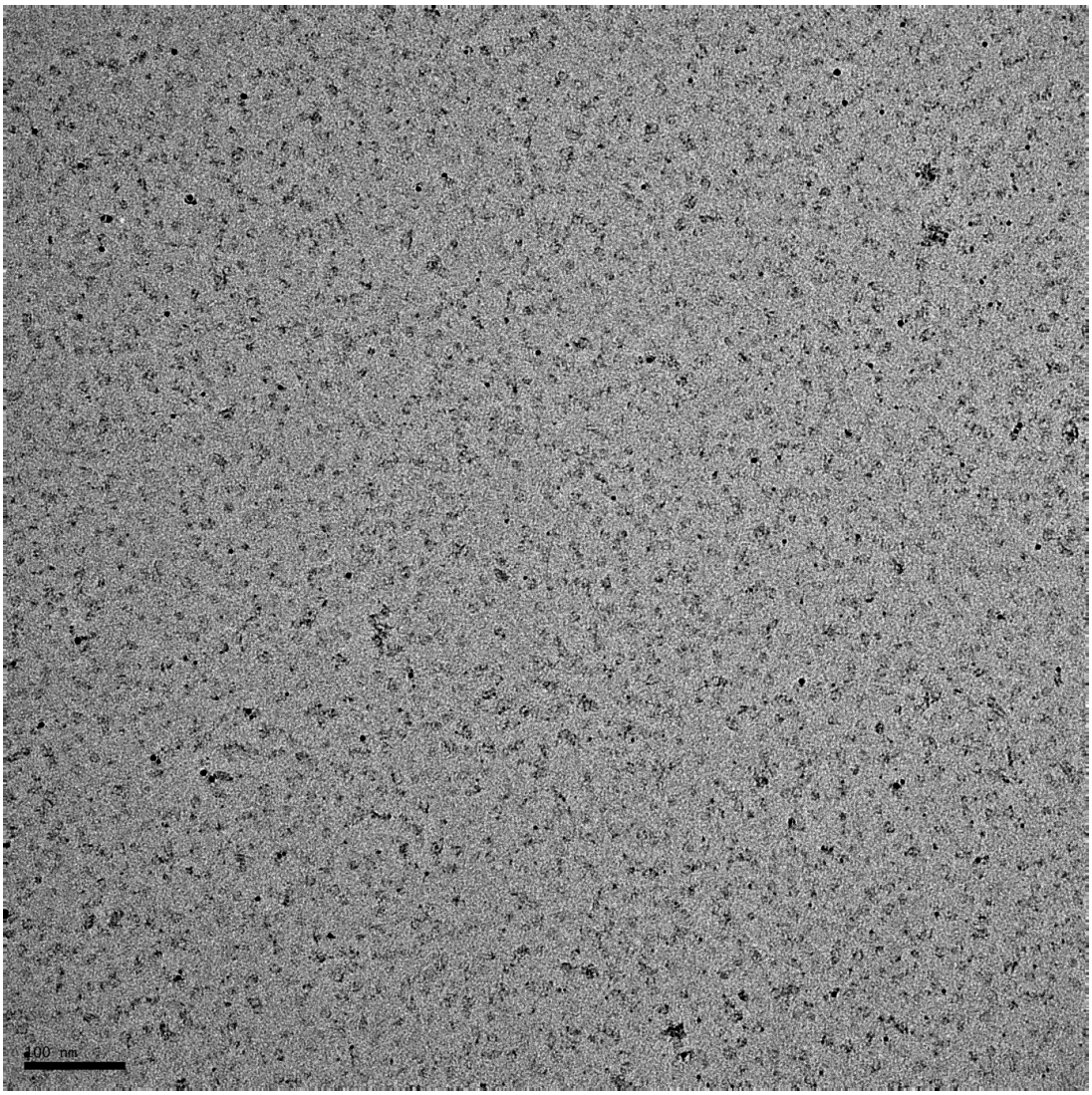



**Figure 3.4.1. Micrograph of 1:8 TatAd-*his* + 1:10 Nanogold + 2x wash**
The micrograph was taken at 57252x magnification under ~1.5 µm defocus. The grid was stained with 2% uranyl acetate. 1.8 nm Ni-NTA-Nanogold® was applied to the grid at ~ 1 nmol/ml, diluted in GF buffer + 0.02% DDM, for 15 minutes, following the 1 minute TatAd sample on-grid incubation. Nanogold appears as intense black dots covering the grid. Scale bar = 100nm

### 3.4.2. The extent of Nanogold labelling is dose-dependent.

An attempt was made to improve overall contrast over the grid whilst demonstrating that the extent of nanogold labelling is dose-dependent. In this case a 1:50 dilution of Ni-NTA-nanogold® was used, and the grid washed only two times with detergent-free GF buffer. The grid was then stained in the same way as before, but using 1% uranyl acetate rather than the 2% solution, as suggested by the Nanogold manufacturer. As such, this grid represents the same washing conditions as the first

labelled sample (Figure 3.4.1) but with a 5x reduced nanogold concentration and 2x reduced stain concentration. An image of this grid is shown in Figure 3.4.2. The number of labelled complexes appears significantly reduced (see arrow heads) whilst the particle boundaries of all complexes are clearly resolved. The morphology of the labelled particles also appears distinct; they are small and round. From these data it can be concluded that the Nanogold is specifically labelling TatAd-*his* complexes however, this dilution does not appear sufficient for optimal labelling of the sample.

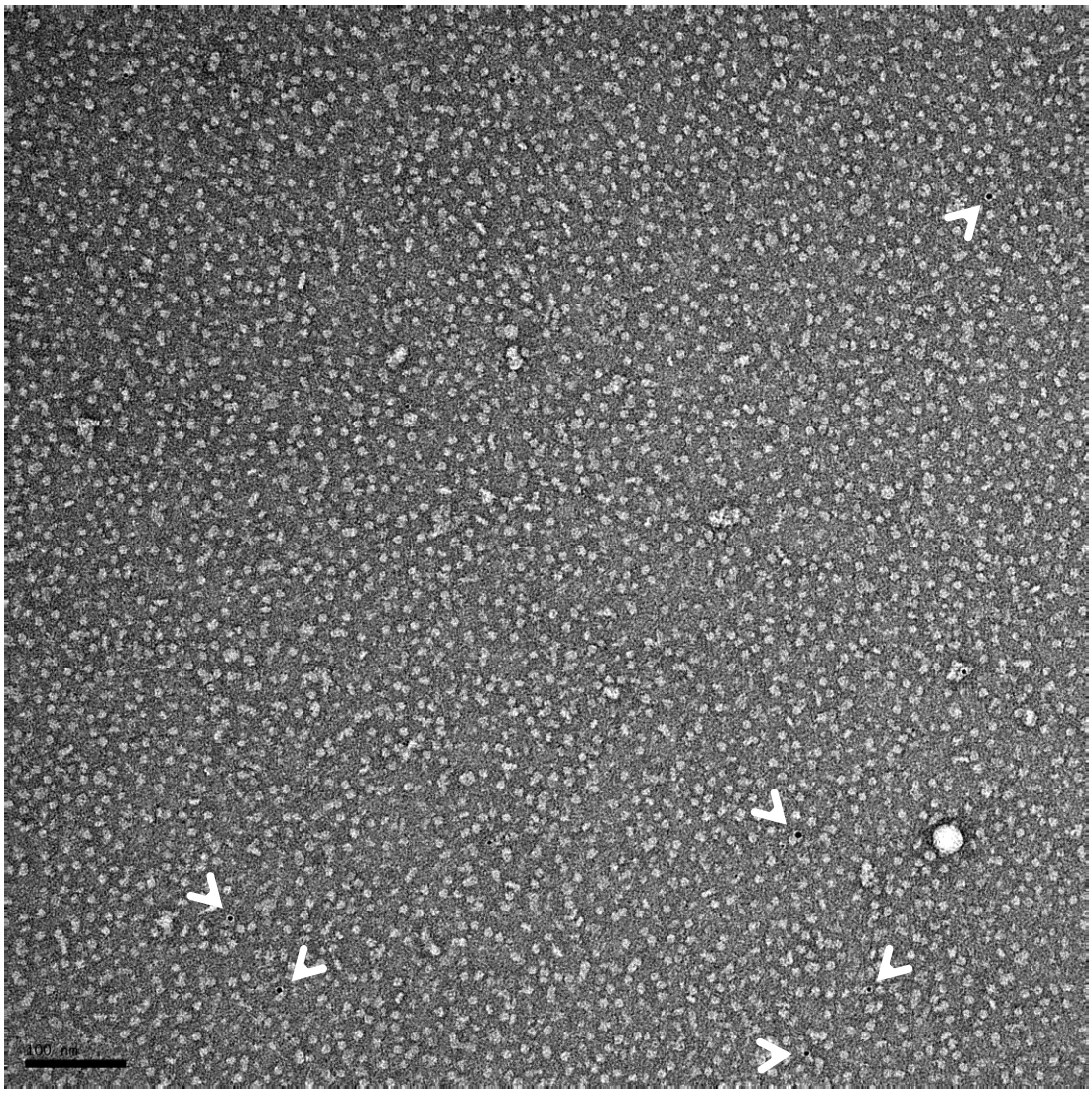



**Figure 3.4.2. Micrograph of 1:8 TatAd-*his* + 1:50 Nanogold + 2x wash**
The micrograph was taken at 57252x magnification under ~1.5 µm defocus. The grid was stained with 1% uranyl acetate. 1.8 nm Ni-NTA-Nanogold® was applied to the grid at ~ 0.2 nmol/ml, diluted in GF buffer + 0.02% DDM, for 15 minutes, following the 1 minute TatAd sample on-grid incubation. Nanogold labelled particle are indicated by white arrow heads. Scale bar = 100nm

### 3.4.3. Effect of extra washing on TatAd-*his* + 1:10 Ni-NTA-Nanogold®

All sample concentrations were kept the same as the first experiment (3.4.1) but the grid was washed an additional two times with detergent-free GF buffer to remove excess Nanogold that may be binding non-specifically. As shown in Figure 3.4.3, the morphology of the particle population has now altered significantly. The grid is now covered in a continuous series of large amorphous aggregates (resembling those shown in Figure 3.3.2) and smaller discrete particles are almost entirely absent. Nanogold particles can still be seen bound to these aggregates indicating that they consist of TatAd-*his*. This suggests that the additional washing of the grid has pushed the DDM concentration below the CMC, forcing the sample to aggregate. This drastic change in particle morphology again demonstrates the effect of the protein to detergent ratio and the minimal contribution of detergent to the normally observed PDCs (as in Figure 3.4.2).

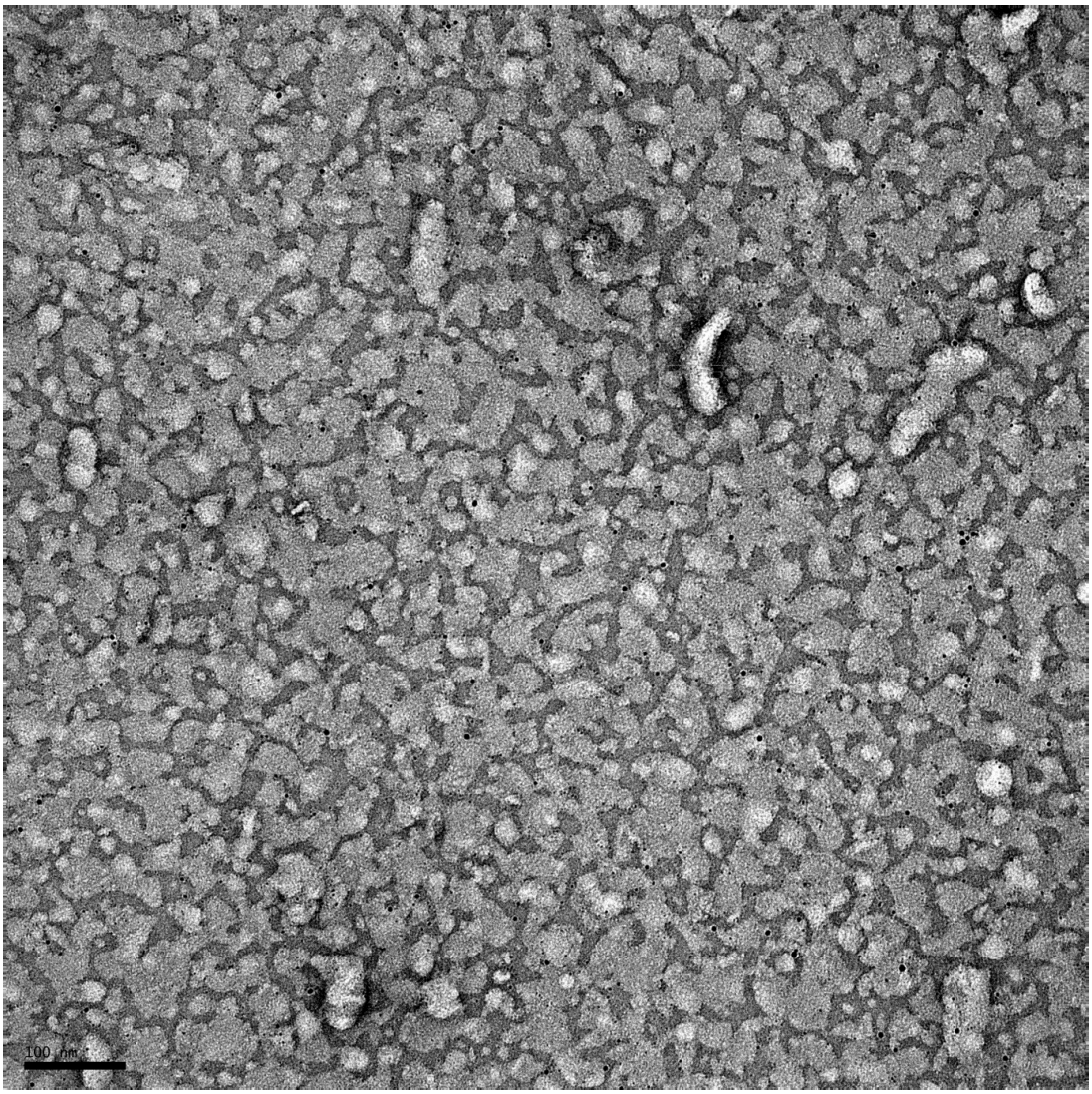



**Figure 3.4.3. Micrograph of 1:8 TatAd-*his* + 1:10 Nanogold + 4x wash**
The micrograph was taken at 57252x magnification under ~1.5 µm defocus. The grid was stained with 2% uranyl acetate. 1.8 nm Ni-NTA-Nanogold® was applied to the grid at ~ 1 nmol/ml, diluted in GF buffer + 0.02% DDM, for 15 minutes, following the 1 minute TatAd sample on-grid incubation. The grid was washed a total of 4 times in detergent- free GF buffer before staining. Nanogold labelled particle are visible as intense black dots. Scale bar = 100 nm

### 3.4.4. Optimised wash method for TatAd-*his* + 1:20 Ni-NTA-Nanogold®

Due to the observed drastic effects of reducing the detergent concentration, a different approach was applied to reduce non-specific Nanogold binding, whilst maintaining TatAd-*his* complex stability, and keeping on-grid detergent levels as low as possible. A 1:20 dilution of Nanogold in GF buffer + 0.02% DDM was used (0.5 nmol/ml), and incubated on the grid following sample addition as before.

Subsequently the grid was washed three times in GF buffer + 0.02% DDM before three washes in the same buffer minus detergent. In this way the grid is being washed a total of 6 times with the washes with detergent compensating for those without. As can be seen in Figure 3.4.4 the particle morphology under these conditions appears much more stable compared to Figure 3.4.3. Distinct, well separated and homogenous particles dominate the grid and overall imaging conditions are much improved compared to Figure 3.4.1. Many labelled complexes are visible, these appear to vary in diameter and in some cases double ringed structures are seen. However, for many of the labelled TatAd-*his* complexes, the Nanogold is visible but the particle boundary with the background is obscured (see selected particles in Figure 3.4.4). This is due to the large difference in contrast between the Nanogold and the protein, compared to the protein and carbon background.

In images taken of more heavily stained areas, the particle boundary for the labelled complexes is much more apparent whereas the resolution of internal density variations in unlabelled complexes is reduced. These features are demonstrated in Figure 3.4.5. Nanogold can be seen binding to small round particles that display a slight size range of ~ 7-9 nm. The Nanogold is bound to the centre of these particles, and the size of the Nanogold cluster increases in proportion to the diameter of the particle, as highlighted in Table 3.4.4.

This suggests that the His-tag, and therefore the C-terminus, of the TatAd-*his* subunits are arranged so as to point towards the middle of the complex. The proportional binding of Nanogold indicates that more His-tags are present in the larger complexes and provides direct visual evidence for the modular assembly of TatAd-*his* complexes. These data support the accepted model of subunit arrangement in TatA-type complex organisation; in which the ‘lid’ of the ring is formed by multiple copies of the amphipathic helices.

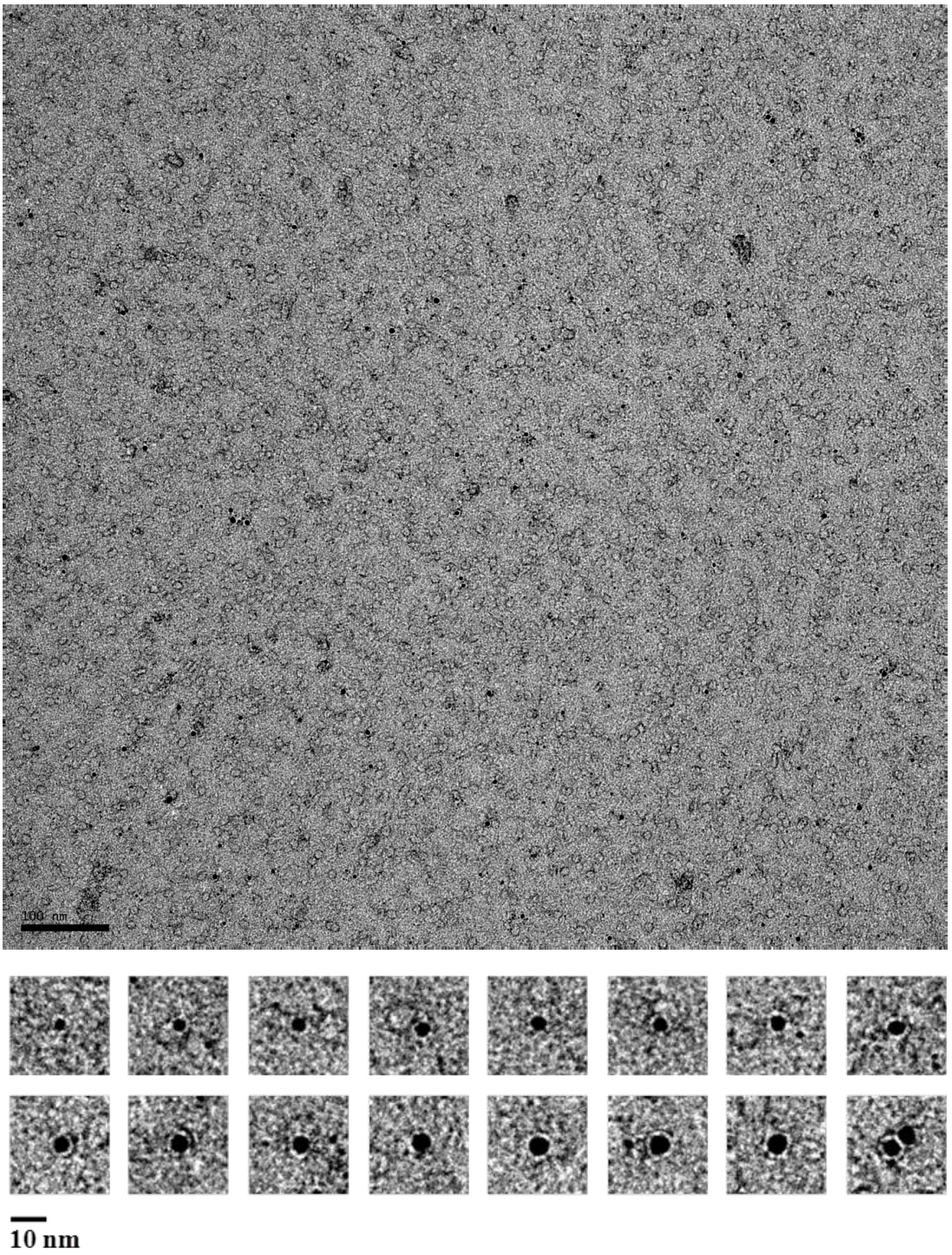



**Figure 3.4.4. 1:8 TatAd-*his* + 1:20 Nanogold, optimised wash, low contrast**
The micrograph was taken at 57252x magnification under ~1.5 µm defocus. The grid was stained with 2% uranyl acetate. 1.8 nm Ni-NTA-Nanogold® was applied to the grid at ~ 0.5 nmol/ml, diluted in GF buffer + 0.02% DDM, for 15 minutes, following the 1 minute TatAd sample on-grid incubation. The grid was washed a total of 6 times; 3 in GF buffer + 0.02% DDM followed by 3 in detergent- free GF buffer, before staining. Nanogold labelled particle are visible as intense black dots Particles shown below were windowed from the micrograph shown. The Scale bar for the micrograph = 100 nm.

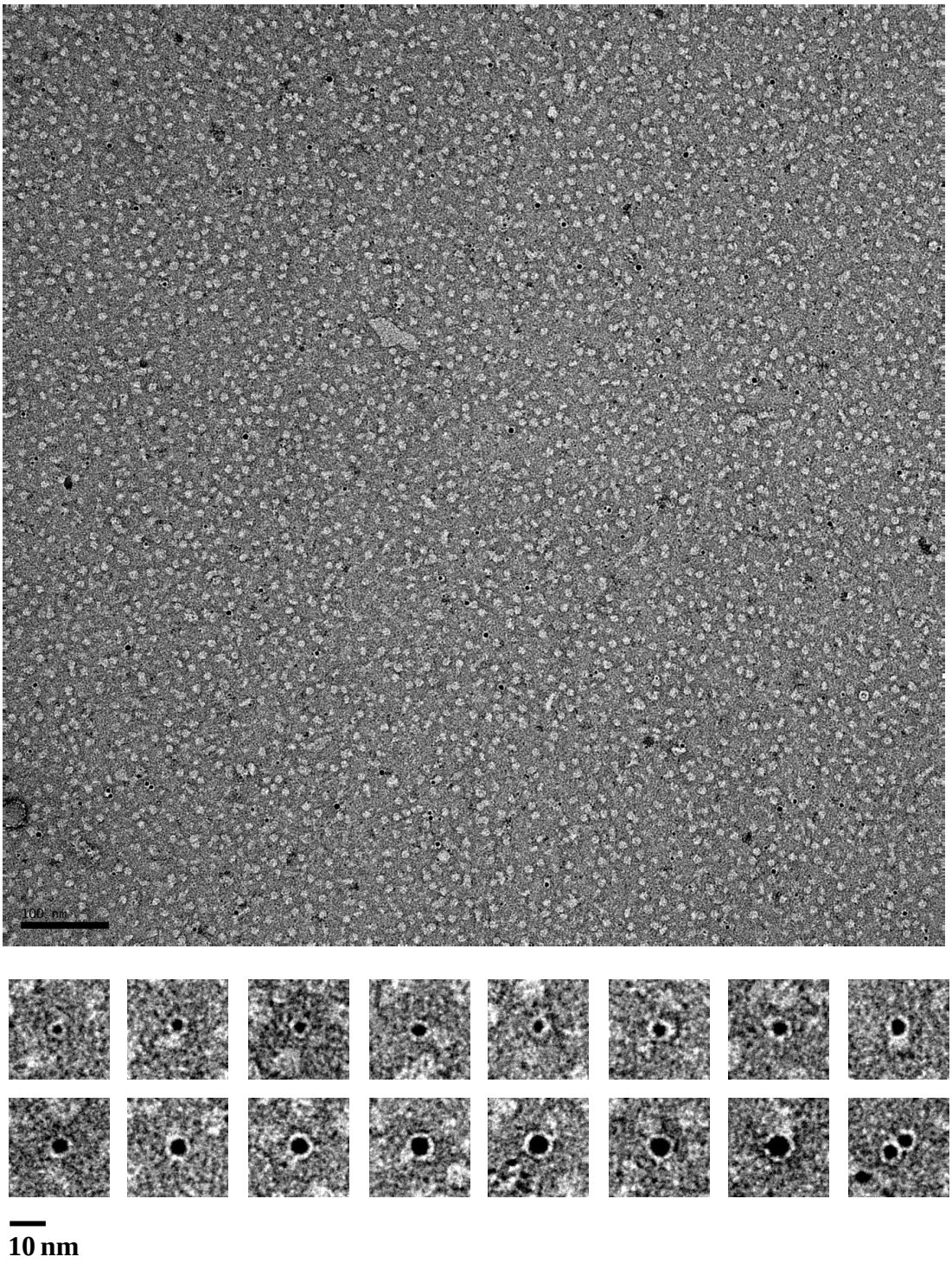



**Figure 3.4.5. 1:8 TatAd-*his* + 1:20 Nanogold, optimised wash, high contrast**
The micrograph was taken at 57252x magnification under ~1.5 µm defocus. The grid was stained with 2% uranyl acetate. 1.8 nm Ni-NTA-Nanogold® was applied to the grid at ~ 0.5 nmol/ml, diluted in GF buffer + 0.02% DDM, for 15 minutes, following the 1 minute TatAd sample on-grid incubation. The grid was washed a total of 6 times; 3 in GF buffer + 0.02% DDM followed by 3 in detergent- free GF buffer, before staining. Nanogold labelled particle are visible as intense black dots Particles shown below were windowed from the micrograph shown. The Scale bar for the micrograph = 100 nm.

| Complex diameter (nm) | Nanogold size (nm) |
|---|---|
| 6.0 | 2.5 |
| 7.5 | 3.6 |
| 8.0 | 4.2 |
| 9.0 | 5.0 |

**Table 3.4.4. TatAd-*his* complex diameter vs. size of bound Nanogold cluster**
The size of the bound Nanogold cluster was seen to increase in proportion with the diameter of the TatAd-*his* complex to which it was bound.

## 3.5. Collecting a tilt series and interactive particle-picking

Having validated the identity of the observed particles the next stage was to proceed towards 3D structural information about the TatAd-*his* complexes. Due to the observed preferential orientation of the particles on the grid, a random conical tilt reconstruction strategy was adopted for image data acquisition. As such, micrographs were collected in tilt-pairs (as shown in Figure 3.5.1.) so as to provide the necessary angular information for the construction of a 3D model.

Micrographs were acquired, first with the microscope stage tilted to 50°, and then at 0°. A total of 133 initial images were taken, comprising 119 tilt-pairs (55 tilted images and 64 untilted). The micrographs were then assessed for drift and astigmatism before estimation of image defocus (see methods chapter 2). Following these quality checks, 22 tilt-pairs of images were selected, covering unique areas of the grid. The average defocus over the untilted micrographs was ~1.5 µm. Next, a total of 2539 initial particle-pairs were windowed out from these selected micrographs into 128x128 pixel boxes using the interactive particle-picking tool in Jweb™.

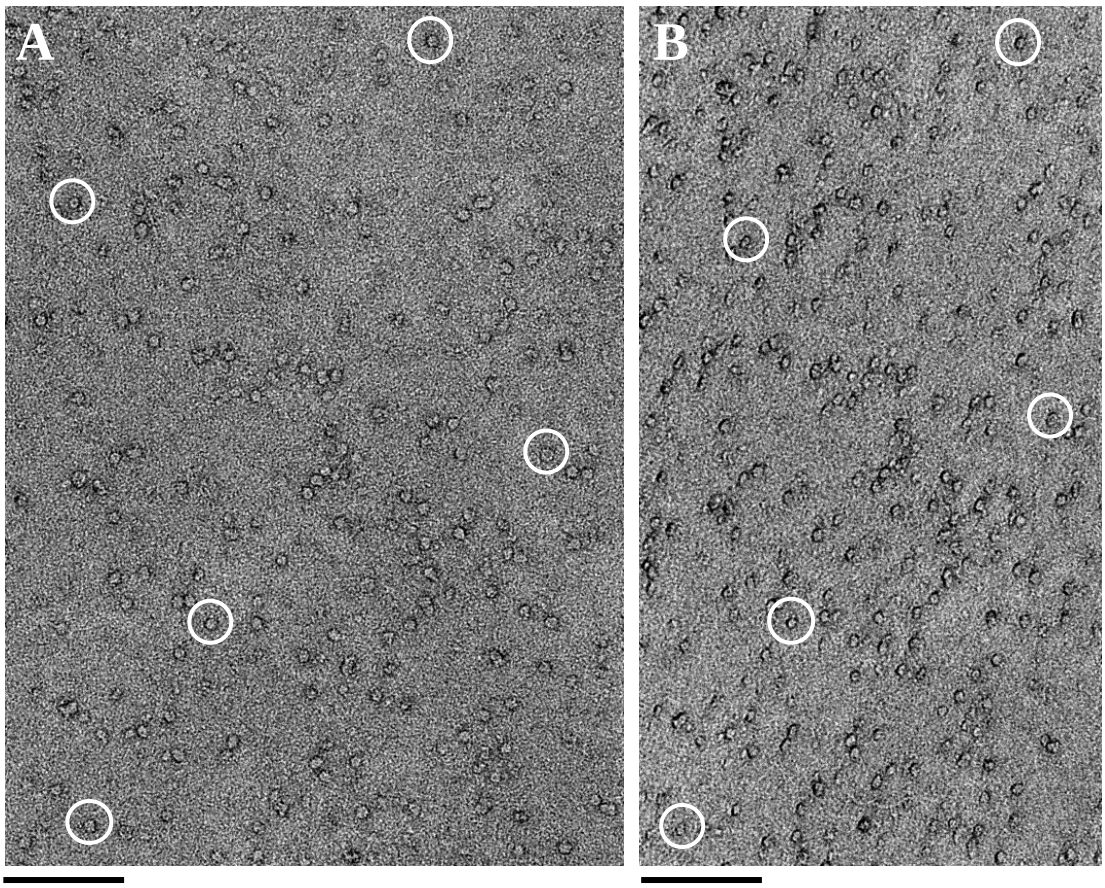



**Figure 3.5.1. TatAd-*his* 1:8 dilution 50° tilt-pair micrographs**
Sections of two micrographs covering the same area of the grid are shown. *A.* microscope stage untilted (0°). *B.* Micrograph stage tilted to 50°. Corresponding particles are circled across both images. The micrographs were taken at 57252x magnification under ~ 1.5 µm defocus. The grid was stained with 2% uranyl acetate. Scale bars = 100nm.

## 3.6. Analysis of potential size variation

From this point the untitled and tilted particle sets were dealt with separately. All 2D classification and refinement was undertaken with the untilted particles only, due to the inherit inability to align the tilted particles in 2D space (see methods). Processing of the tilted particles is covered later in the chapter.

### 3.6.1. Initial processing of particles for downstream alignment and classification

The untitled particles were subjected to a manual, visual assessment, using Python program *montagefromdoc*, to interactively edit the data-set. Broken and anomalous particles were discarded, as well of those where the windowed area clipped the edge of the original micrograph (see Figure 3.6.1). As shown in Figure 3.6.2 an initial, crude, reference-free rotational and translational alignment of all the particles revealed a small, round average of ~ 8 nm diameter; consisting of several globular densities, with a central pool of stain indicating a cavity or channel in the structure. The edited particle set of 2438 particles were then processed and enhanced via several processes in preparation for further analysis, the effect of these processes are shown in Figure 3.6.3. First all pixel values were normalised, to remove negative values, before the images were corrected for the effect of the contrast transfer function (CTF), applied during image acquisition. CTF-correction was performed using the phase-flipping method based on the defocus calculation for each micrograph. Finally, a low-pass Butterworth filter was applied to the particles to remove high resolution noise from the images. With the pass and stop bands set to 0.175 and 0.5 respectively, all information lower than ~ 15 Å resolution was preserved while information above ~ 5 Å was removed, with a smooth transition between.

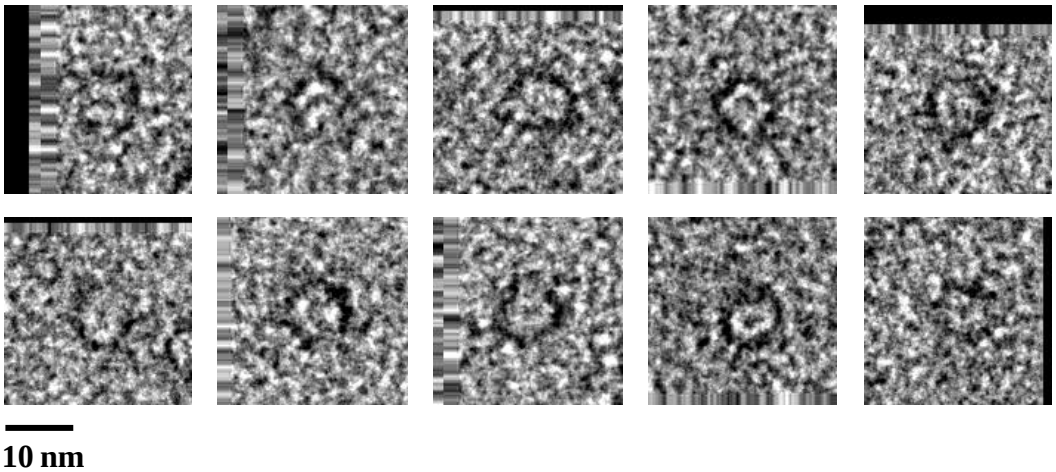



**Figure 3.6.1. Manually rejected TatAd-*his* particles**
Clipped particles were rejected from the initial data-set due to the anomalous pixel fringe present around the edge of the CCD acquired micrographs. These values interfere with the various alignment methods used during image processing.

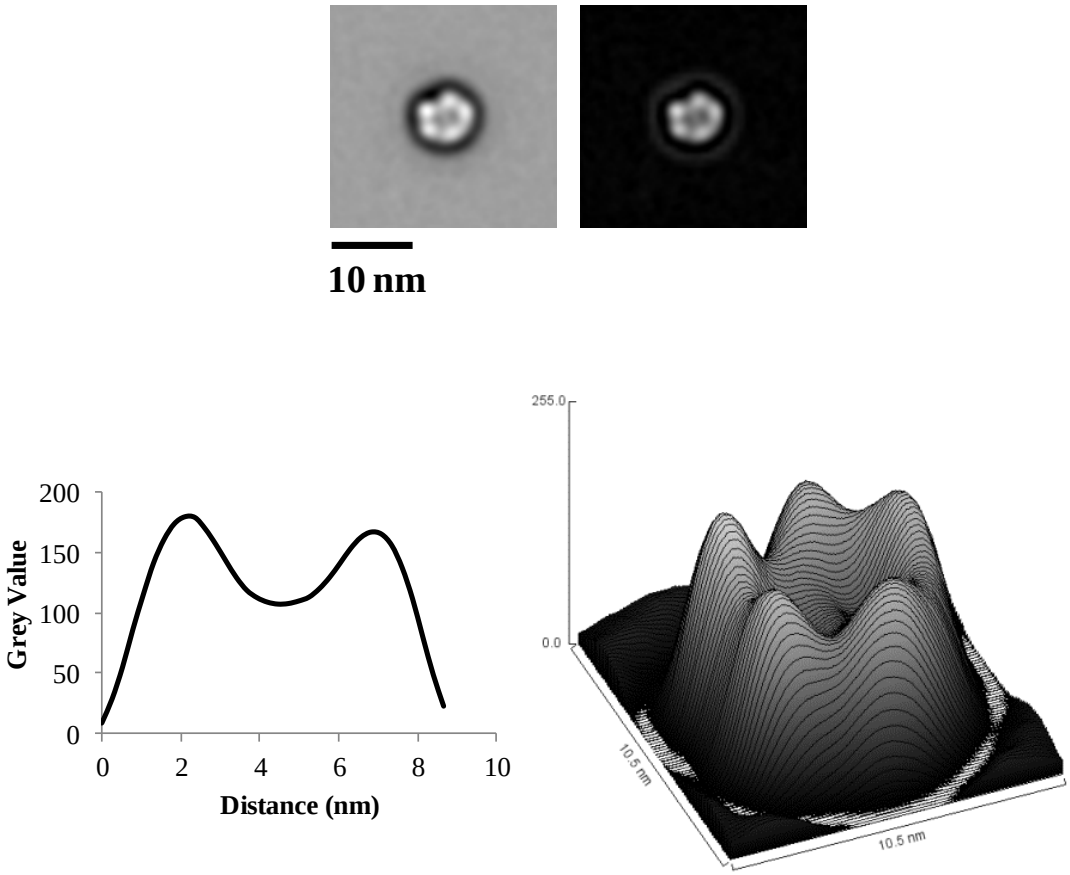



**Figure 3.6.2. Crude initial alignment of TatAd-*his* particles**
All particles were aligned rotationally and translationally to reveal a ring shaped structure consisting of multiple globular densities (top left, a background subtracted image is shown on the right). 2D and 3D profiles across the aligned average image display a central cavity, as indicated by the reduced pixel density (grey value), with the 3D profile also showing the globular ring densities. Profiles produced using ImageJ (ver. 1.44p) *plot profile* and *surface plot* tools.

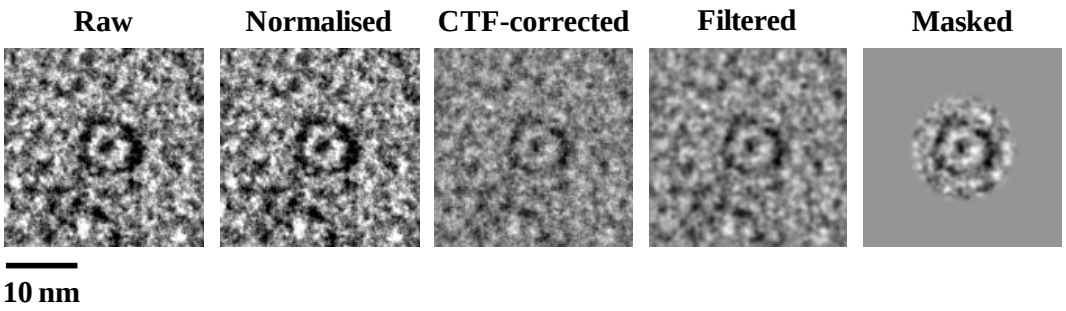



**Figure 3.6.3. Processing of TatAd-*his* particles to restore image information**
The same particle image is shown after the successive application of each processing step. During normalisation all negative pixel values are removed. CTF-correction involves compensating for a mathematically applied contrast enhancing function applied during the image acquisition (see methods section). Filtering removes high resolution noise that is reducing the S/N ratio. Masking eliminates background features in the image from interfering with the downstream alignment steps (the mask shown is for illustrative purposes only; a mask with a smoother Gaussian edge distribution was used during processing).

### 3.6.2. Initial centring and size classification

Particles were subjected to 3 rounds of iterative centring, to account for manual inaccuracy in particle windowing, and to remove the effect of positional bias on the estimation of size variance, as shown in Figure 3.6.4. A multivariate statistical analysis (MSA) using a correspondence method was then performed on the centred particles to examine the most significant eigenfactors (see methods section).

In Figure 3.6.5 the top 10 factors of image variance are displayed as montages of both importance images and reconstituted average particles. Here the characteristic concentric ring pattern indicating size variance (White *et al.*, 2004) is displayed in factor 5; in this case indicating that size variation is not the most significant factor of pixel variance within the data-set. However, when a wider radius mask (34 pixels) is applied to the particles, size variance now appears as the top factor. This indicates that the initial mask was too tight and obscured some of the larger particles from the analysis.

Particles were then separated based on this eigenfactor via a hierarchical ascendant classification using Ward's linkage method, as shown in Figure 3.6.6. A suitable truncation point was selected from the dendrogram produced and class averages were

generated for each of the 6 size classes. Visual analysis of the constituent particles in the class averages indicated that size variance was still present in some of the classes along with some anomalous and off-centre particles. In light of this further editing of the data-set was deemed necessary to improve the classification.

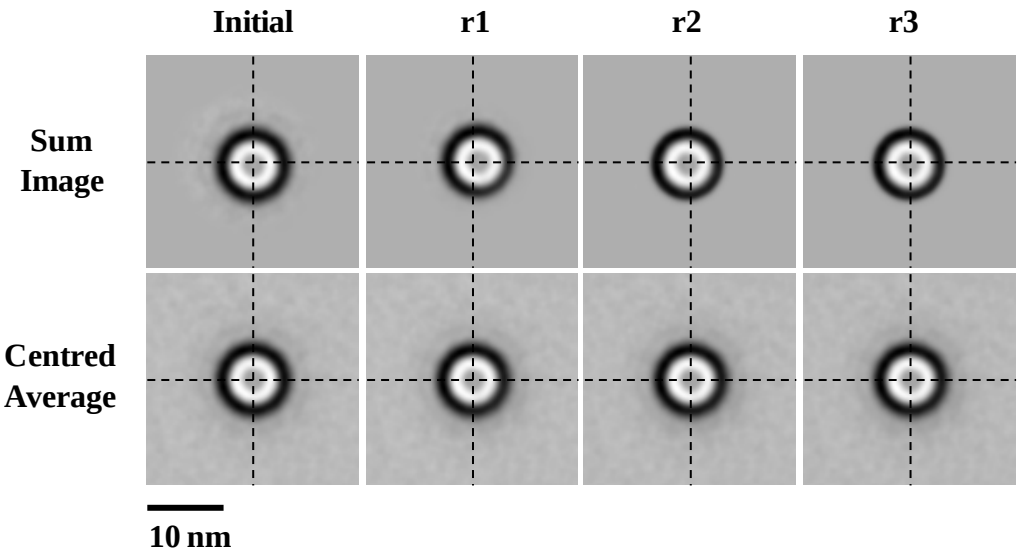



**Figure 3.6.4. Initial iterative centring of TatAd-*his* complexes**
Initial centring was performed using a 34 pixel radius mask with a 6.0 pixel Gaussian half-width. This was reduced to 25 pixels for the first round of iteration and 20 pixels for the subsequent 2 rounds. Particles were aligned only translationally. The degree of centring was measured using the centre of gravity (CG) SPIDER command. The results given as x,y coordinates were as follows: R1 = -0.1,-0.07, R2 = -0.03, 0.03, R3 = 0.01, 0.02 (0.0, 0.0 = centre of image). Diameter of final average = ~ 7.8 nm.

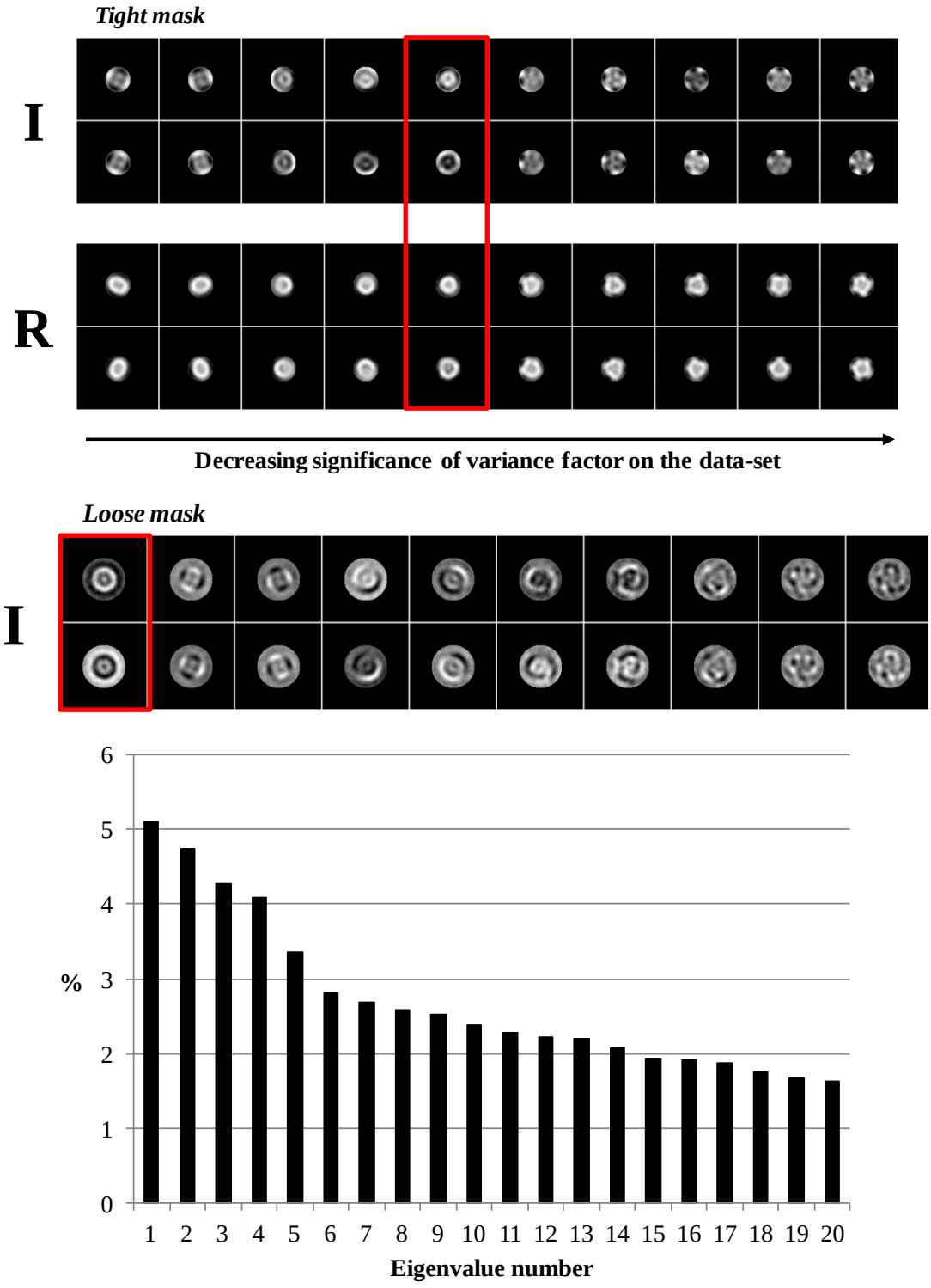



**Figure 3.6.5. Initial correspondence analysis of TatAd-*his* complexes**
Using a tight initial mask size variation appears as the 5$^{th}$ higher factor in pixel variance within the data-set (red box), as indicated by the concentric ring pattern in the importance images (I) and the small and large ringed particles shown in the reconstituted image pair (R). A second attempt using a 34 pixel radius mask displays a concentric ring pattern indicative of size variation in the top eigenimage (highlighted in red). The graph shows the impact of each factor on image variance in the data-set. The top 5 factors have the most significant impact within the images.

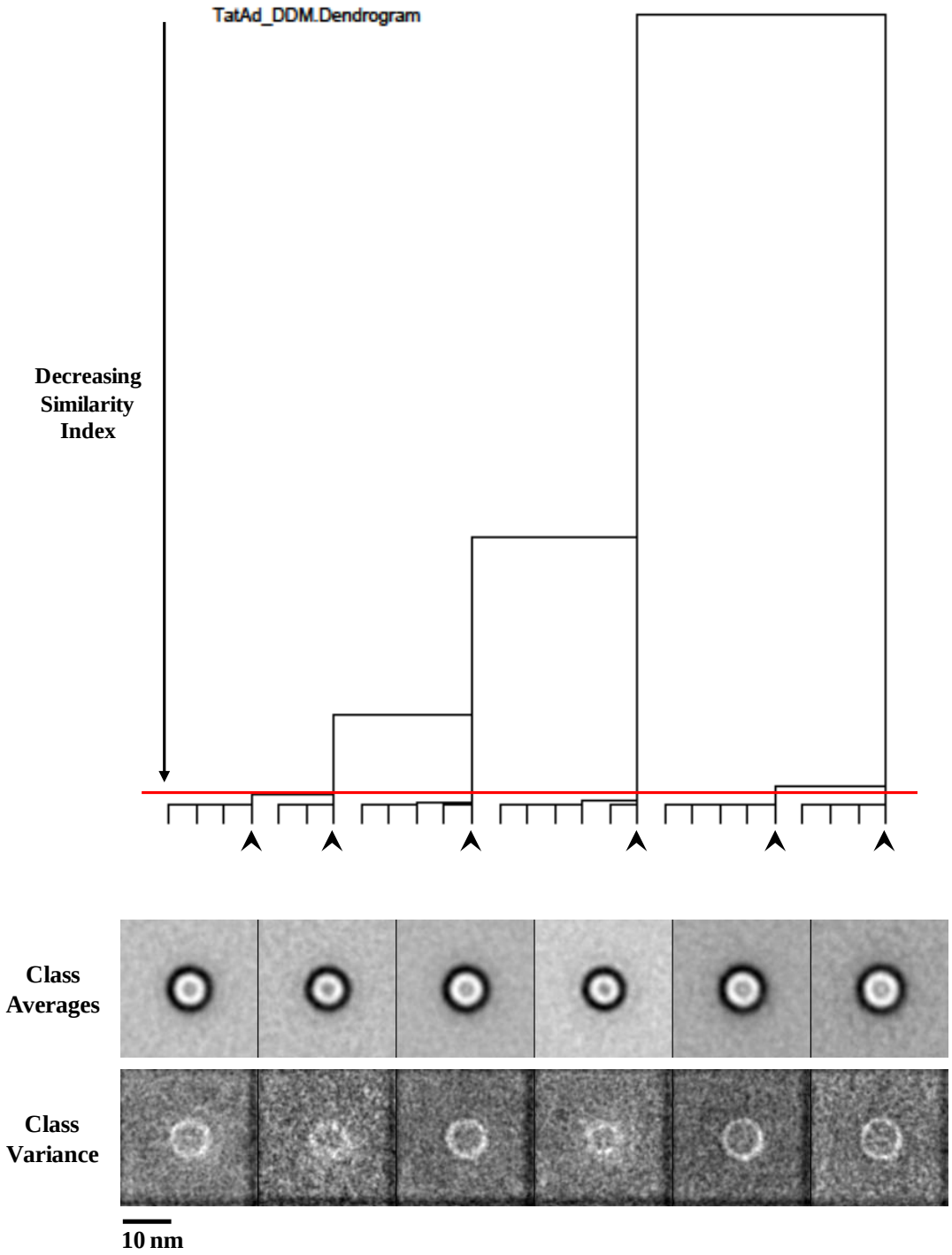
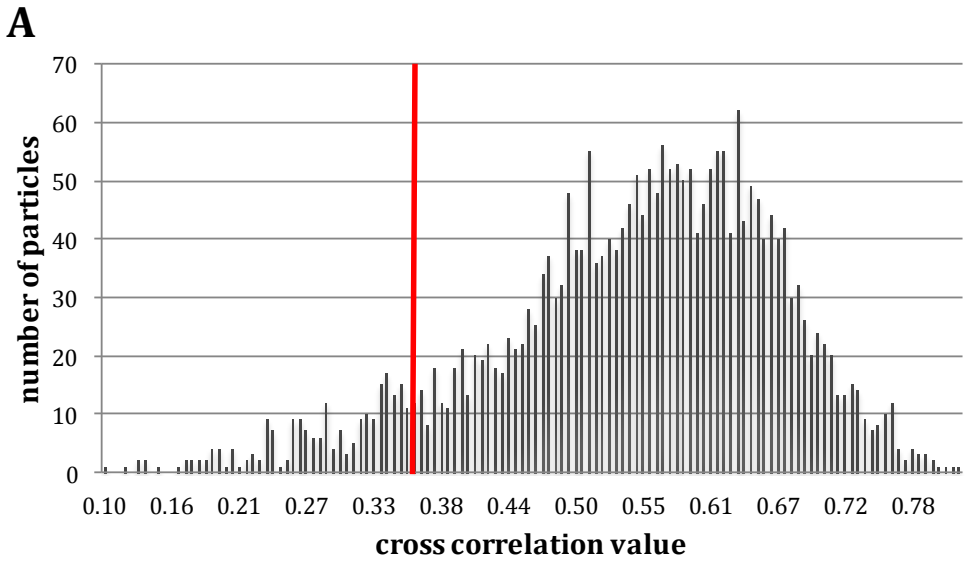



**Figure 3.6.6. Initial size classification of TatAd-*his* complexes**
HAC was performed based on the top eigenfactor (see Figure 3.6.5.) using Ward's method of separation. This resulted in the classification as shown in the dendrogram. The lowest possible cut-off value was selected (see red line) to give the maximum number of classes while maintaining a significant step-change in similarity index between branching points; in this case resulting in 6 significant classes (indicated by arrowheads). Branching below the selected cut-off is shown for illustrative purposes only. Average images are shown for each class displayed in the same order (from left to right) as the corresponding dendrogram branch points. Variance images are shown below, where areas of high pixel variance across the class are displayed in white.

### 3.6.3. Editing of the data-set using an iterative cross-correlation method

All particles were cross-correlated to the total centred sum image to assess gross similarity to the average particle morphology. As shown in Figure 3.6.7.A. the majority of particles show a high correlation to this average image, whilst a minority show little to no correlation at all. These anomalous particles were removed to improve the S/N ratio and allow for meaningful classification of the data.

The 10% of particles with the lowest cross-correlation value were selected and trimmed from the data-set as shown. The remaining particles were re-centred and this process was repeated a total of 3 times, in each case the lowest 10% of particles were removed, resulting in a final edited data-set of 1912 particles. Using this iterative approach provides all particles with an opportunity to re-align to the average and find a better fit before the similarity check and subsequent editing. As shown in Figure 3.6.7.C. all particles with a CCC < 0.4 have been removed by the 3$^{rd}$ round of editing, this indicates that anomalous particles with very little similarity to the average are no longer included in the data-set.



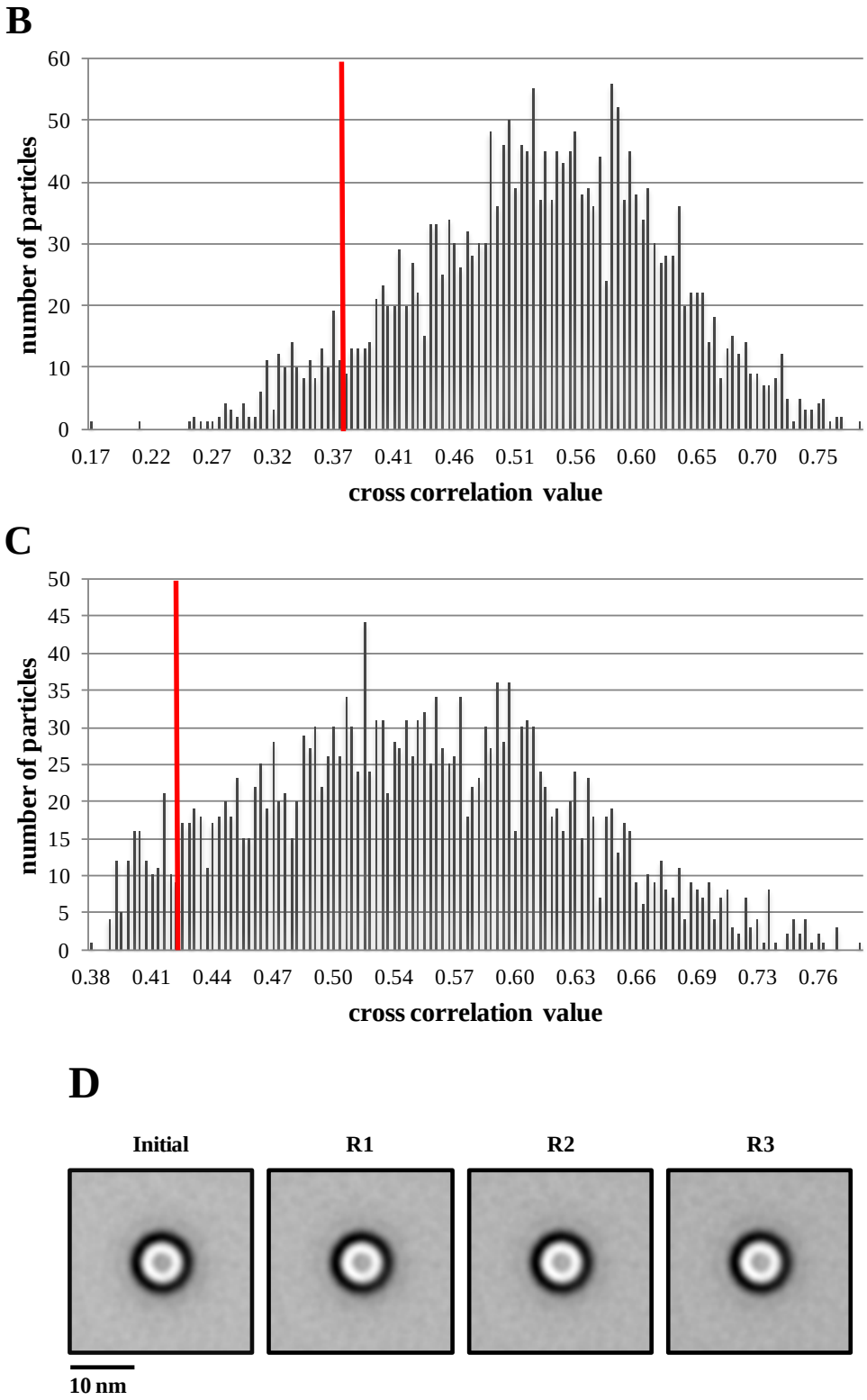
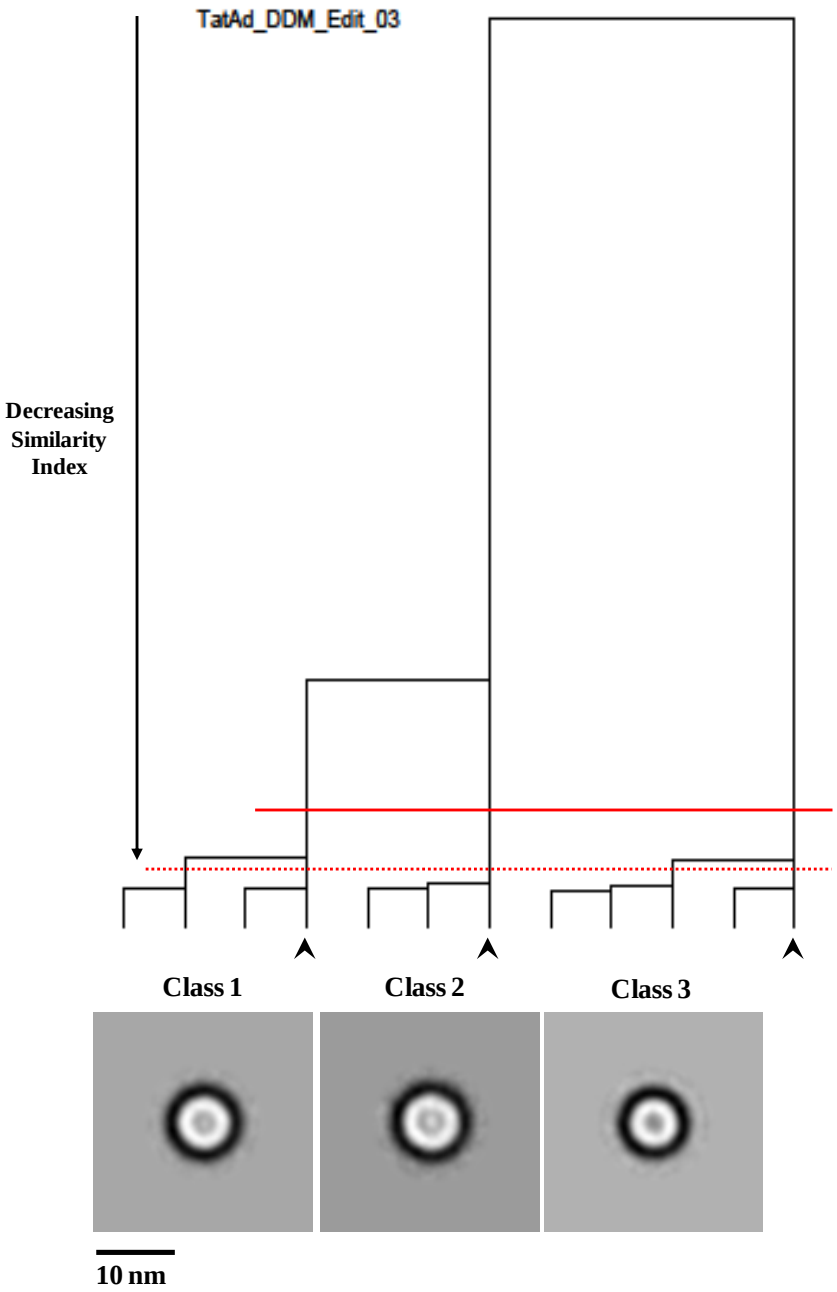



**Figure 3.6.7. Refinement of TatAd-*his* data-set by a cross-correlation method**
*A.* All particles were cross-correlated to the total centred sum image. Similarity is measured as a value between 0 and 1, with 1 meaning total similarity, and the results given as a histogram. The bottom 10% of the particles were removed from the data-set, the cut-off is indicated by the red line. The remaining particles were re-centred and the process repeated twice (*B* + *C*). *D.* The average sum image is not affected by this process.

### 3.6.4. Size classification of the edited data-set

Classification of the edited data-set resulted in 3 significant classes based on Ward's method of HAC. As can be seen from Figure 3.6.8 these classes show a variation in diameter of ~ 1 nm although the visual difference between the averages is quite clear. The medial sized Class 1 is the most populous, whereas the largest sized Class 2 is the least and displays central density occluding the 'pore' seen in other averages. A further branching into 5 classes (dashed-red line) was discarded because it led to an increase in the standard deviation of pixel variance.

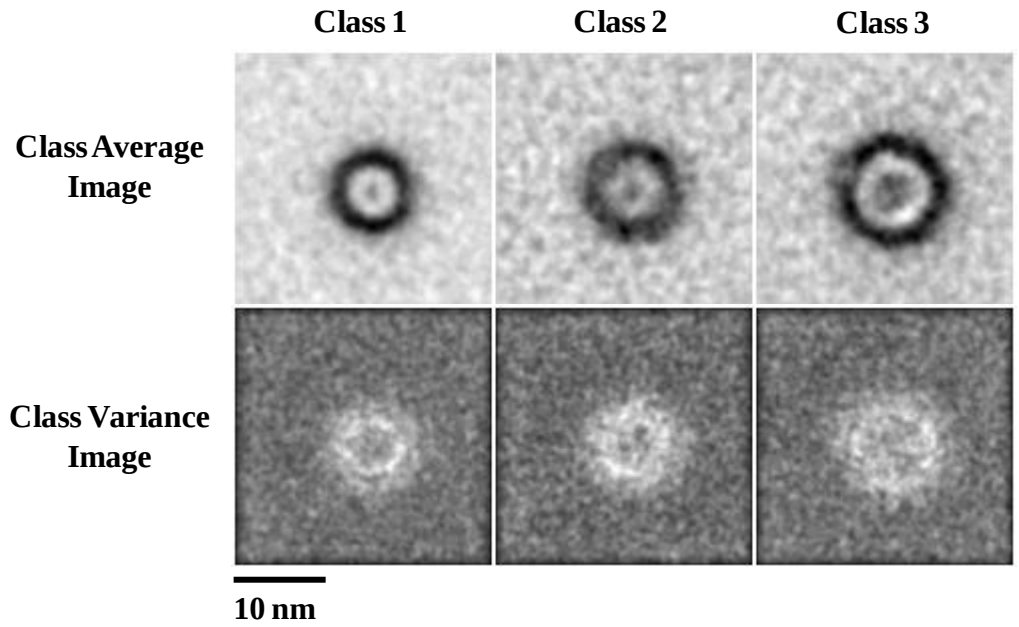


| **Class #** | **1** | | **2** | **3** | |
|---|---|---|---|---|---|
| **Diameter (nm)** | **7.9** | | **8.4** | **7.3** | |
| **# particles per class** | **856** | | **382** | **674** | |
| ***# particles per class*** | ***488*** | ***368*** | ***382*** | ***490*** | ***184*** |
| **Central density?** | **Slight** | | **Yes** | **No** | |
| **SDVAR** | **0.0725** | | **0.0892** | **0.0727** | |
| ***SDVAR*** | ***0.0760*** | ***0.0865*** | ***0.0892*** | ***0.0757*** | ***0.0971*** |

**Figure 3.6.8. Size classification of the refined TatAd-*his* complex data-set**
HAC was performed based on the top eigenfactor only, using Ward's method of separation. This resulted in the classification as shown in the dendrogram. The selected cut-off value is shown as a red line (maximum index = 2840, cut-off = 300). This was chosen to give the maximum number of classes while maintaining a significant step-change in similarity index between branching points; in this case resulting in 3 significant classes (indicated by arrowheads). Average images are shown for each class displayed in the same order (from left to right) as the corresponding dendrogram branch points. The dashed red line represents a cut-off producing 5 classes, in this case classes 1 and 3 both sub-classified. This further classification was discarded due to its negative impact on the standard deviation of variance as shown in the table (italicised).

### 3.6.5. Rejected particles

The rejected pool of 616 particles were analysed to check for mistakenly discarded particles and to look for any consistent features of interest. Centring of these particles proved problematic due to the large disparity in appearance and extent of contrast; therefore multivariate statistical analysis (MSA) was performed on the un-centred particles. Hierarchical ascendant classification (HAC) was then conducted as before, with the selected cut-off resulting in 3 significant classes displaying a very clear size variation as shown in Figure 3.6.9.



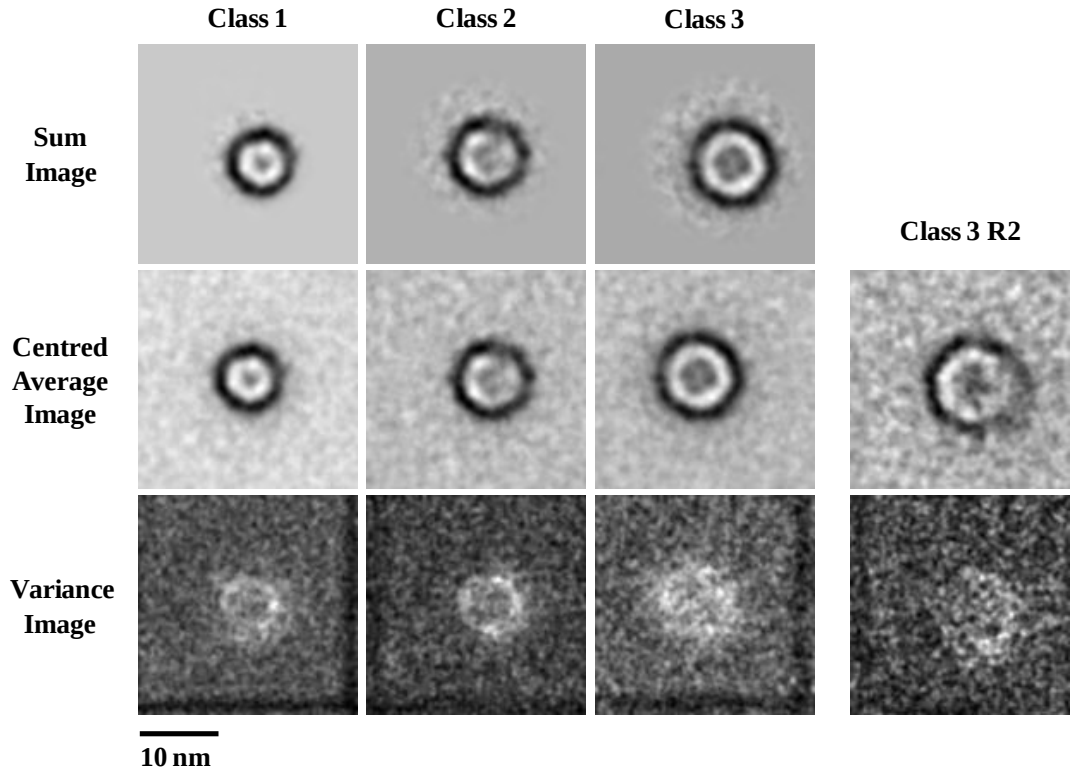

| Class # | 1 | 2 | 3 |
|---|---|---|---|
| # particles per class | 253 | 168 | 195 |
| SDVAR | 0.152 | 0.177 | 0.171 |

**Figure 3.6.9. Size classification of rejected particle set**
Class averages were generated by HAC on the rejected pool of 616 particles, based on the top eigenfactor, representing size variance. The class variance images display a high degree of pixel variation (white pixels) across the entirety of the averages and the surrounding space; this indicated a heterogeneous particle population. The table below indicates the distribution of rejected particles and the standard deviation of pixel variance (SDVAR) within each class.

As can be seen from Figure 3.6.9 these class averages are blurred and show a high degree of pixel variance across their entire surface indicating a low degree of similarity within the classes. This is also shown by the high standard deviation of pixel variance. Due to the heterogeneity within the classes, the particles do not centre well and have low cross-correlation to their respective class average.

After classification the particles were centred, resulting in the averages shown in Figure 3.6.10. The class 1 average of the rejected particle was seen to resemble that of class 3 of the selected particles (see Figure 3.6.8). Therefore these particles were cross-correlated against the centred average of selected class 3, with a manual cut-off set at 0.4 CCC, so as not to increase the overall variance within the class. This resulted in 78 extra particles being added to selected class 3.

Manual analysis of the constituent particles revealed classes 2 and 3 to be of poor quality however, consisting of mostly broken or low contrast particles as indicated by the variance images in Figure 3.6.10. From class 3 only 29 particles were identified with a diameter of over 9 nm but when these were re-centred (Class 3 R2) the resulting average is spurious. Manual analysis of these particles (Class 3 R2) showed them to vary considerably in size and shape (see Figure 3.6.11), and as such they were discarded.




| Class # | 1 | 2 | 3 |
|---|---|---|---|
| Diameter (nm) | 6.8 | 9.0 | 10.0 |

**Figure 3.6.10. Centring of rejected particle classes**
The class averages (sum images) were centred using the standard method resulting in the centred average images. The class variance images display a high degree of pixel variation (white pixels) across the averages and the surrounding space; this indicated a heterogeneous particle population. Class 3 R2 is a manually edited version of class 3 consisting of the 29 largest particles. The table below indicates the diameter of each class average.

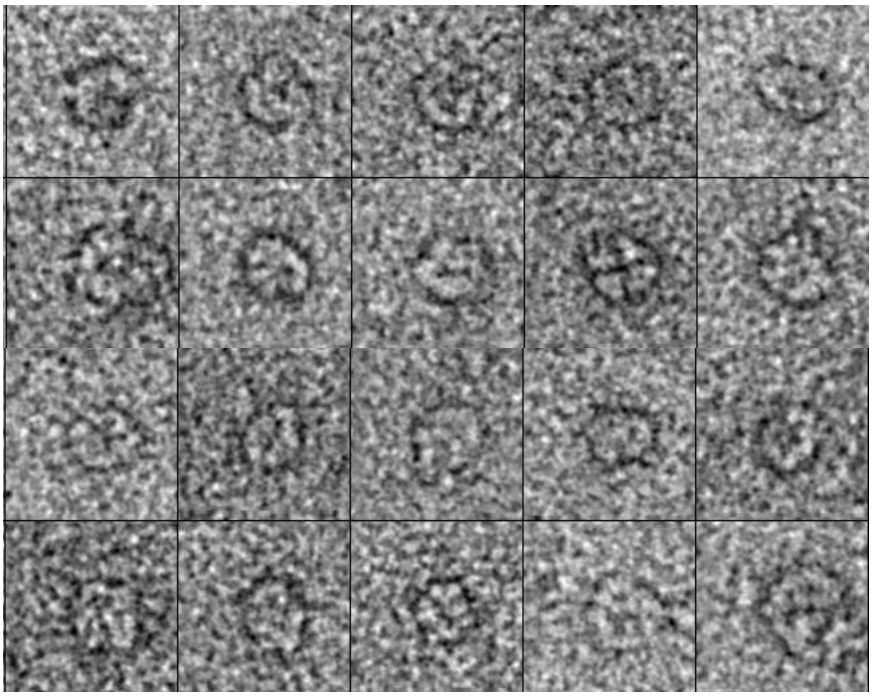
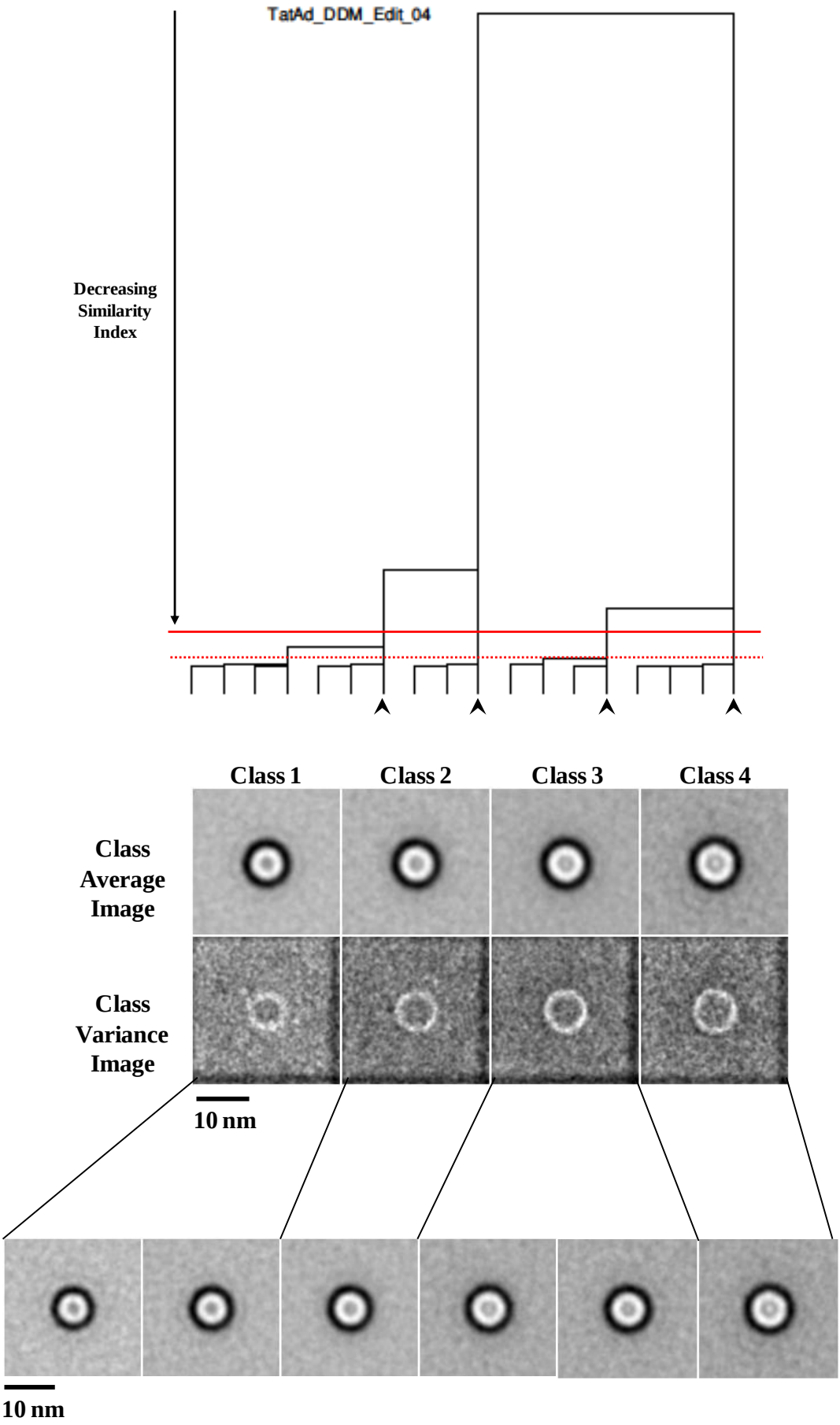



**Figure 3.6.11 Montage of particles from rejected class 3**
A selection of particles from class 3 R2 are shown, indicating the variable size and shape of the constituents of this rejected class.

### 3.6.6. Selection of a final size classification and subsequent multi-reference based re-assignment

The 1990 selected particles were re-classified, following the classification method discussed previously. A final cutting level producing 4 size classes was selected as shown in Figure 3.6.12. A lower cut-off producing 6 size classes was rejected as this revealed no further size separation and simply increased the pixel variance within each class.

In order to further validate this classification, a multi-reference refinement method was applied. The centred class averages were used as cross-correlation references for the whole data-set, and the particles were re-classified based on the most similar reference. This method allows inter-class particle movement, to determine the most appropriate assignment for each particle.


TatAd_DDM_Edit_04
Decreasing Similarity Index
Class 1
Class 2
Class 3
Class 4
Class Average Image
Class Variance Image
10 nm
10 nm

| **Class #** | **1** | | **2** | **3** | | **4** |
|---|---|---|---|---|---|---|
| **Diameter (nm)** | **7.5** | | **7.9** | **8.5** | | **9.0** |
| **# particles per class** | **586** | | **500** | **641** | | **263** |
| ***# particles per class*** | ***232*** | ***354*** | ***500*** | ***330*** | ***311*** | ***263*** |
| **Central density?** | **No** | | **No** | **Slight** | | **Yes** |
| **SDVAR** | **0.0755** | | **0.0791** | **0.0765** | | **0.0953** |
| ***SDVAR*** | ***0.0921*** | ***0.0819*** | ***0.0791*** | ***0.0865*** | ***0.0865*** | ***0.0953*** |

**Figure 3.6.12 Selection of the final size classification**
A final round of HAC was performed on the particle set of 1990 particles. At the selected cut-off level (red line) this resulted in the 4 size classes shown. A further

Figure 3.6.13.A. displays the results of the first round of reassignment. The total height of each bar represents the number of particles within the class prior to reassignment, and the colour coding represents the distribution of particles after class reassignment. So the sum of the blue bars indicates the total number of particles in the reassigned class 1. As can be seen, the majority of the particles remain within their initially assigned class, indicating that this classification is valid, and those that do move, shift proportionately to the size differences between adjacent classes. As such, a small number of particles initially assigned to class 3, move to classes 2 + 4, but hardly any move to class 1. However, the stability of the classes is uneven, with class 1 the most stable and class 4 the most unstable and also least populous. After this re-classification the particles were again centred within each class and the multi-reference refinement was repeated; in this second round the shuffling between classes was much less pronounced as shown in Figure 3.6.13.B. All 4 classes appear stable with roughly equal residual inter-class particle movement. The final distribution of particles between these size classes shows the smallest class (class 1) to be the most abundant, and the largest (class 4) to be the least populous; with almost twice as many particles in class 1 than class 4.

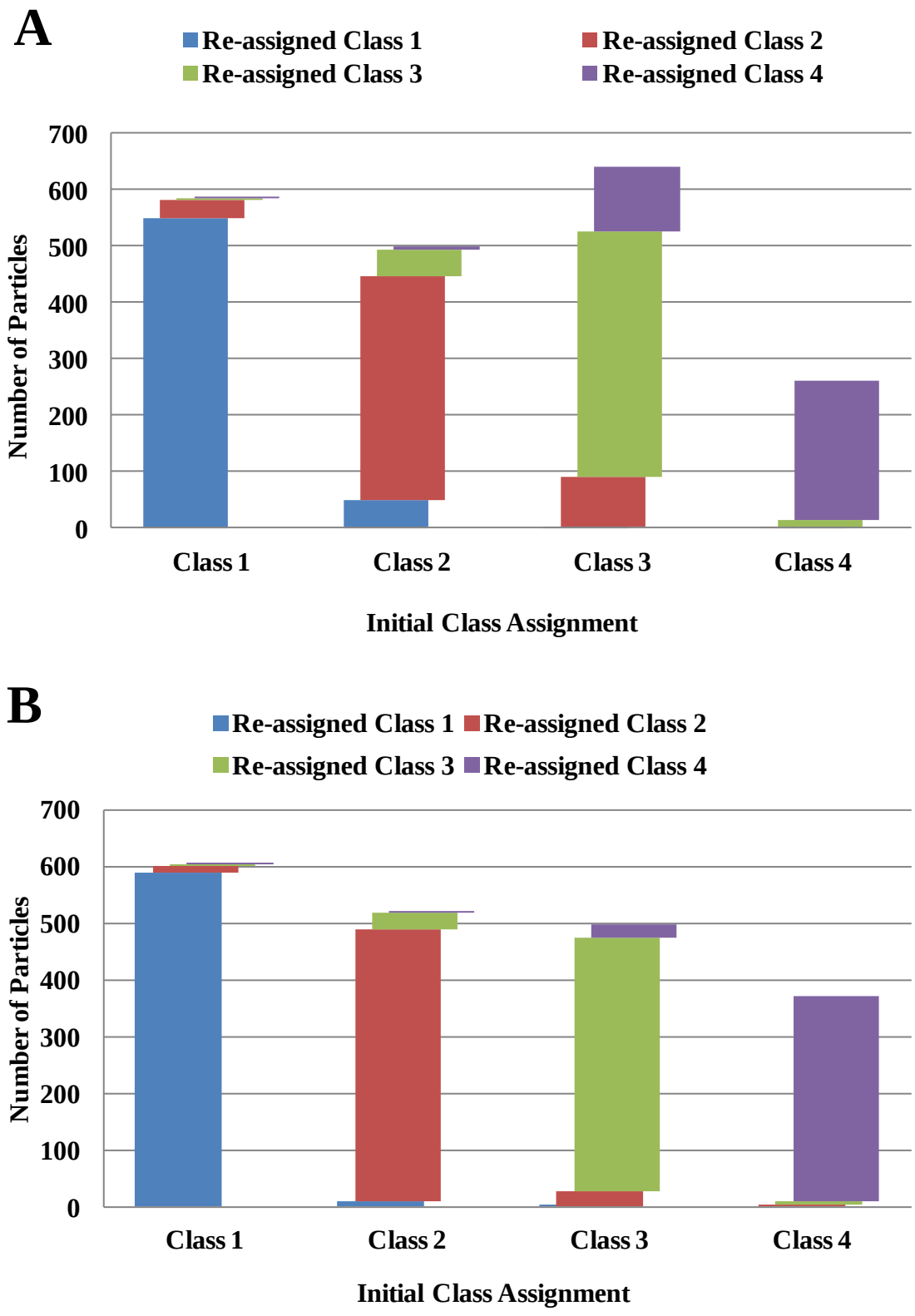
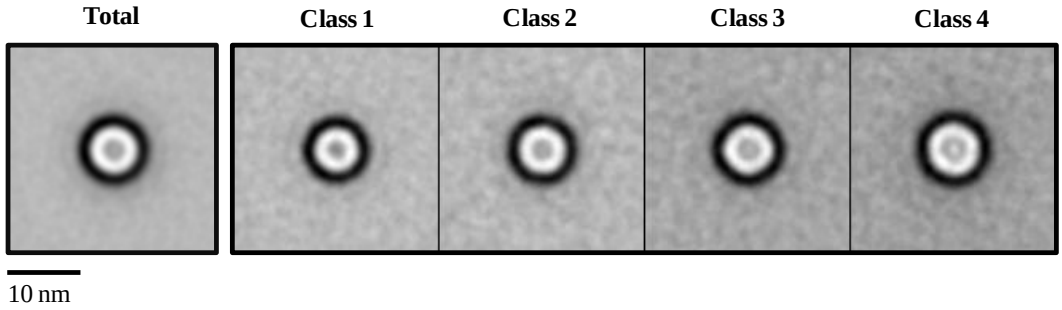



**Figure 3.6.13. Iterative class re-assignment of particles during multi-reference refinement of TatAd-*his* complex size classification**

*A.* All particles were cross-correlated against the 4 size class averages and reassigned to whichever class they showed the most similarity. *B.* All particles were then re-centred within their new class and the process repeated. The total height of each bar represents the number of particles within the class prior to reassignment. The colour coding represents the distribution of particles after class reassignment.

The class averages after the final centring procedures are shown in Figure 3.6.14 along with the diameter and particle counts. The estimated size variance between classes is only ~ 1.5 nm unlike the 4.0 nm difference reported previously for TatA complexes (Gohlke *et al.*, 2005). Interestingly, the presence of central density was also seen to correlate with particle size. In the class 1 average a clear central pool of stain can be seen, whereas in the class 4 average, white density is seen to obstruct this cavity. All classes are well represented with enough constituent particles to be statistically relevant. The size of the particles is inversely proportional to their abundance within the data-set.




| Class # | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| Diameter (nm) | 7.5 | 7.9 | 8.5 | 9 |
| # particles | 586 | 500 | 641 | 263 |
| # particles post-APMS R1 | 604 | 519 | 496 | 371 |
| # particles post-APMS R2 | 598 | 521 | 485 | 386 |

**Figure 3.6.14. Results of TatAd-*his* complex multi-reference refinement**
Translationally aligned final size class averages are shown along with the unclassified total average for reference. The class averages increase in diameter from class 1-4 with a total variation of ~ 1.5 nm. Particle counts for each class are provided prior to the re-classification procedure, after the first round (R1) and after the second round (R2). APMS = SPIDER operation for multi-reference based translational (x, y shift) alignment.

### 3.6.7. Rotational alignment of size-separated particle classes

After the particles were classified based on size, each class was processed separately. For each class, a reference-free translational and rotational alignment was applied to the original un-aligned particles. In this way, there is no bias towards a reference, no compounding of previous alignment errors, and the particles are only aligned with respect to others in the same size class. This program was run repeatedly as a different random starting particle is used each time, to which all other particles are aligned. As such, the final alignment differs slightly each time, but no major differences were observed between runs.

The final aligned class averages are shown in Figure 3.6.15. Each average appears as an asymmetrical ring consisting of globular densities each measuring ~ 2.5 nm across; with 6 densities present in the largest class average (class 4) and 5 in the others (classes 1-3). No significant variation in the thickness of the ring could be detected between the classes. However, the central cavity varies slightly in diameter; from ~ 2.5 nm for class 1, to ~ 4.2 nm for class 4, in the longest dimension. The intensity of the central density can also be seen to clearly vary across the class averages, as highlighted in the density profiles shown in Figure 3.6.16.

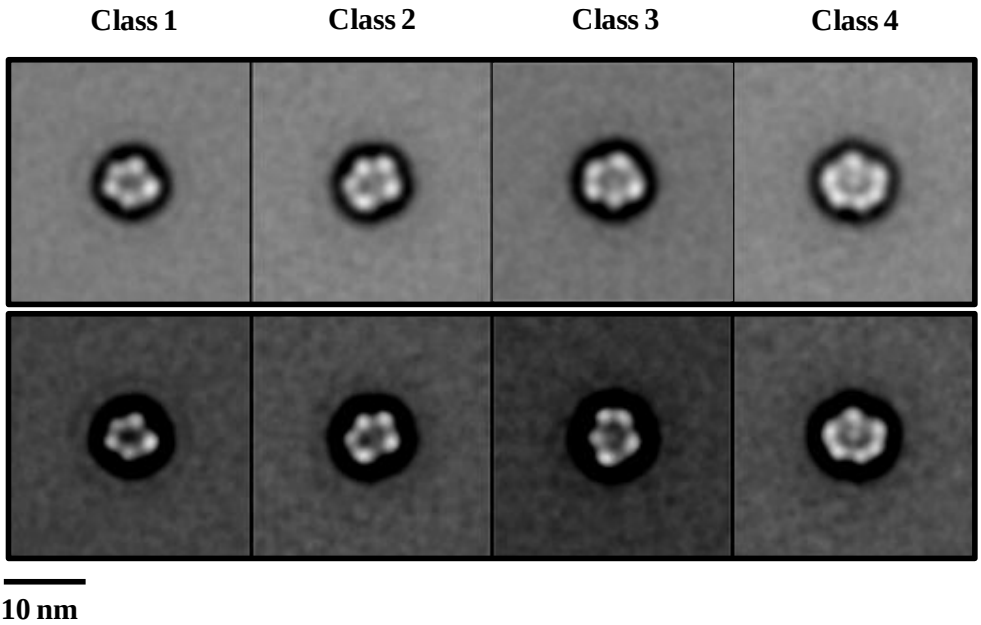



**Figure 3.6.15. Rotationally aligned TatAd-*his* class averages**
Reference-free rotational and translational alignment was applied to the size separated class averages (shown in Figure 3.6.14). Lower panel is displayed with reduced brightness to enhance the globular ring densities.

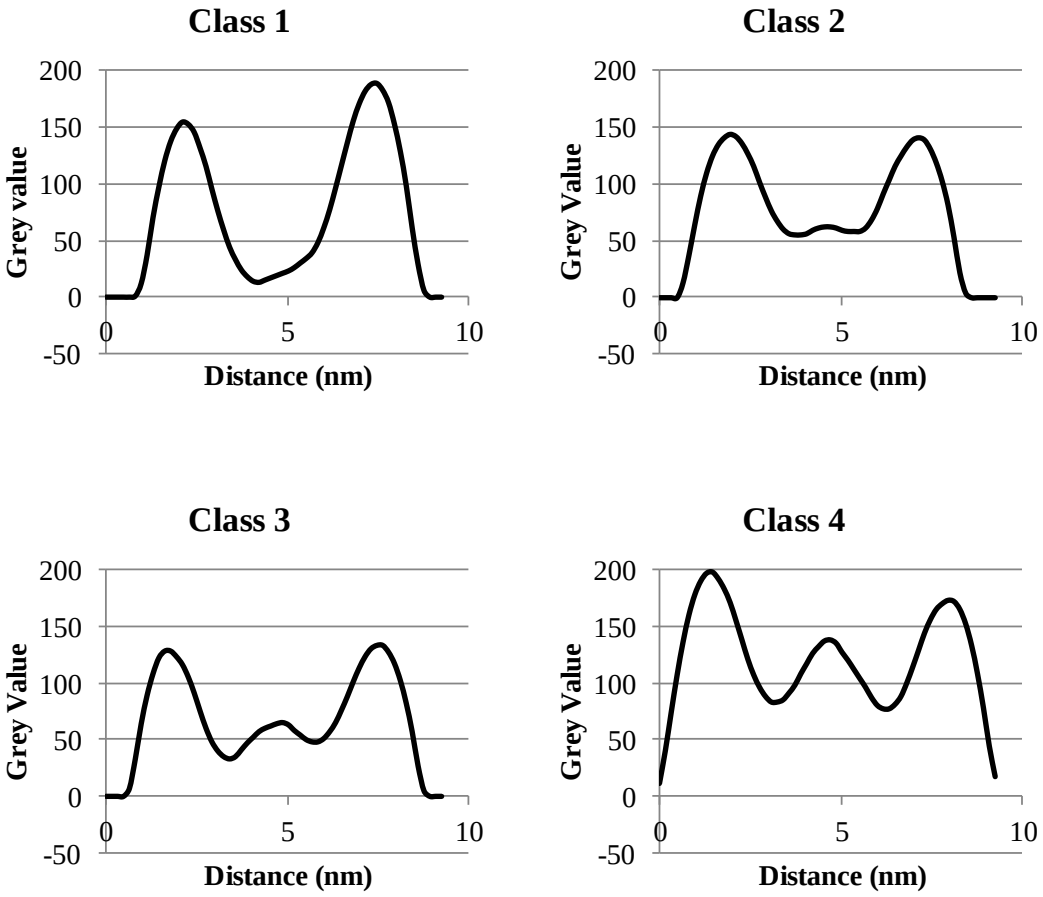



**Figure 3.6.16. Comparison of TatAd-*his* class average density plots**
For each class a 1D density profile was made across the rotationally aligned average image. For class 1 a central cavity is indicated by the reduced pixel density (grey value) in the middle of the plot. In classes 2 and 3 the relative density change between the ring and central cavity is less pronounced but still clear. In class 4 a clear peak of density is present within the central cavity indicating an obstruction of some kind. Profiles were produced using the ImageJ (ver. 1.44p) *plot profile* tool.

## 3.7 Processing of the tilted particles

### 3.7.1. Initial processing of the tilted particle data-set

The tilted particles were normalised as for the untilted particles and CTF-corrected using the same overall phase-flipping method. However, due to the tilt applied to the micrographs the defocus of each tilted particle is dependent upon its position relative to the tilt axis. Therefore a different program, CTFTILT (Mindell and Grigorieff, 2003), was used to determine the defocus range across the grid and the defocus of each particle was calculated using its corresponding image coordinates (see methods section for explanation). Every tilted particle was then CTF-corrected individually based on the defocus of the image. No filtering was applied to the tilted particles at this point.

### 3.7.2. Custom-masked method for centring of the tilted particles

In order for the rotational alignment of each class to be accurately applied in three-dimensional space, the tilted particles must be appropriately centred. Therefore, modifications were made to the centring method used on the untilted particles to apply to the oval-shaped tilted particles. The 50° tilt used on this data-set introduced an artefact into the tilted particles. In Figure 3.7.1 an extra fringe of density can be seen in the top left on the tilted sum image (indicated by a red arrowhead). This is due to the exaggeration of the contrast difference between the stain around the particle and the carbon background in the direction of the applied tilt. With a standard circular mask this fringe was seen to interfere with the centring procedure with detrimental effects on the 3D models generated later. To avoid this problem an iterative custom masking and centring procedure was implemented. As shown in Figure 3.7.1 the particles are moved off-centre using the original method (top right panel) but over a series of iterations the particles are accurately centred using the custom masked approach regardless of the particle size.

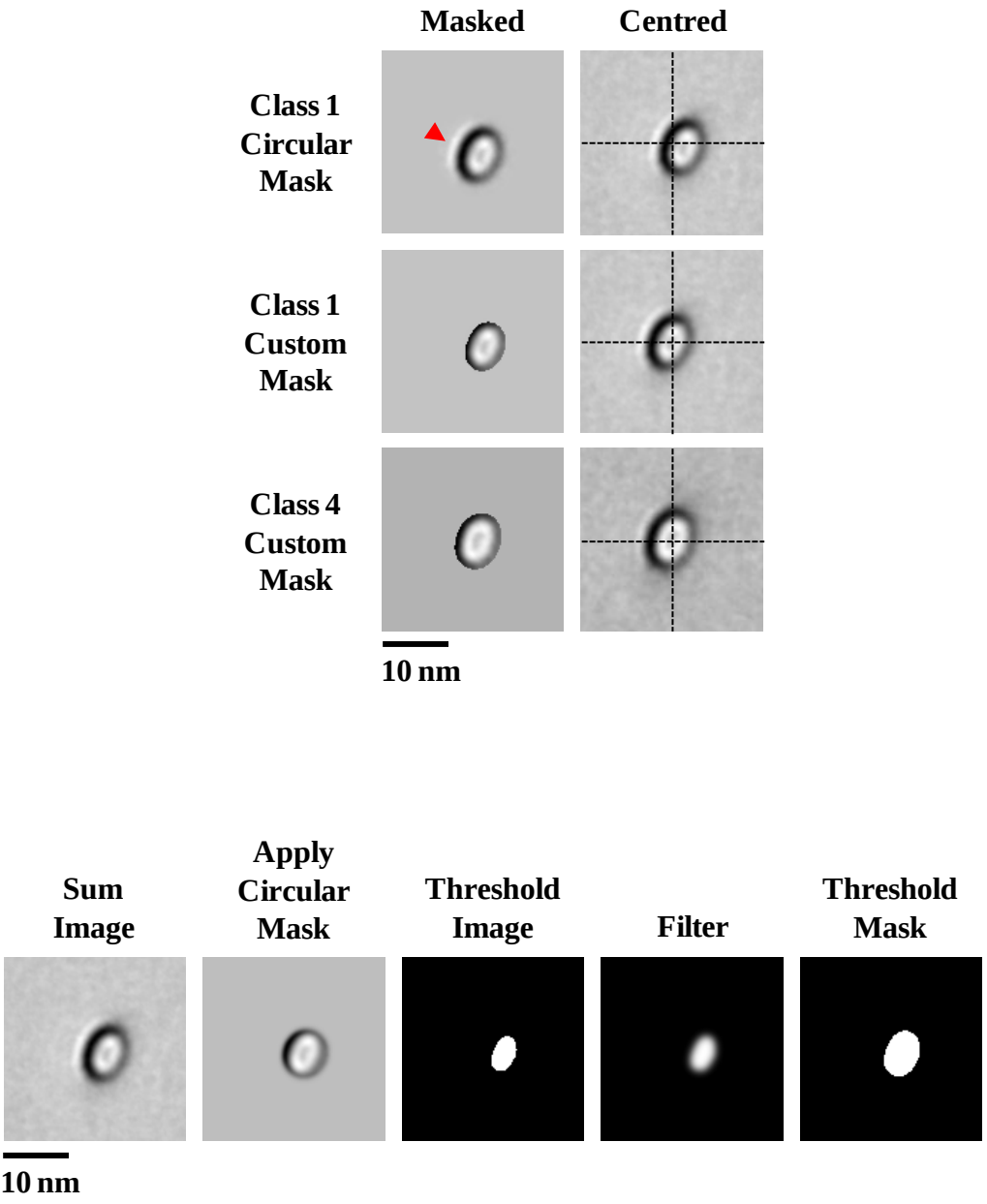



**Figure 3.7.1. Custom centring method for tilted particles**
*Top.* Using a standard circular mask for the centring of the tilted data-set (top left panel) results in the particles appearing off-centre (top right panel). This is due to the impact of the observed density fringe (indicated by the red arrow head). With the application of a custom mask that traces the outline of the particle this problem is rectified regardless of the size of the particle. *Bottom.* The custom mask is generated by first applying a tight circular mask to the sum image. A binary image is then created by applying a threshold to the masked image. This is then filtered to provide a smooth edge and finally another threshold is applied to create a binary mask of greater diameter.

## 3.8. 3D structural models of TatAd complexes

For each class the required angles to produce a 3D model were collated into an Euler angle document file: θ (Known tilt angle), φ (In-plane rotation from untilted particles) and ψ (default set to zero).

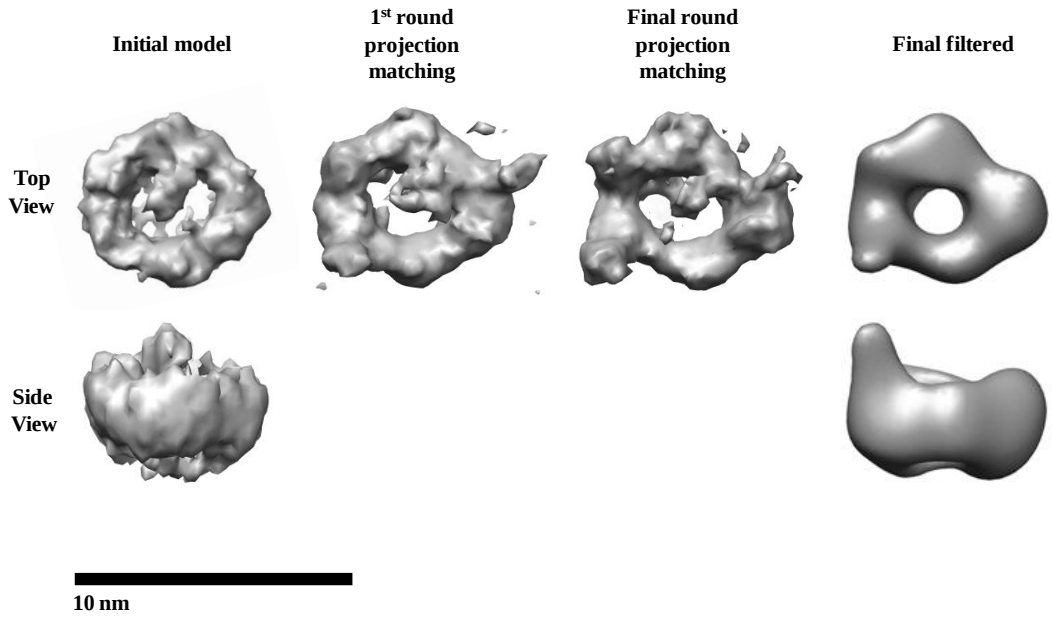
Initial 3D reconstructions were produced for each size class using the faster but less accurate direct Fourier back-projection method (BP 32F).

These structures were then refined using a projection matching method. In this way each particle is allowed to find a more appropriate fit within the model with the overall aim of improving the final resolution. 2D projections of the initial models were generated at regular angular increments covering a full rotation (0°-359°), and a tilt angle (θ) of between 45° and 50° to account for any localised variation of the carbon film across the surface of the grid. The particles were then cross-correlated to each projection in turn before re-alignment based on the best matching projection.

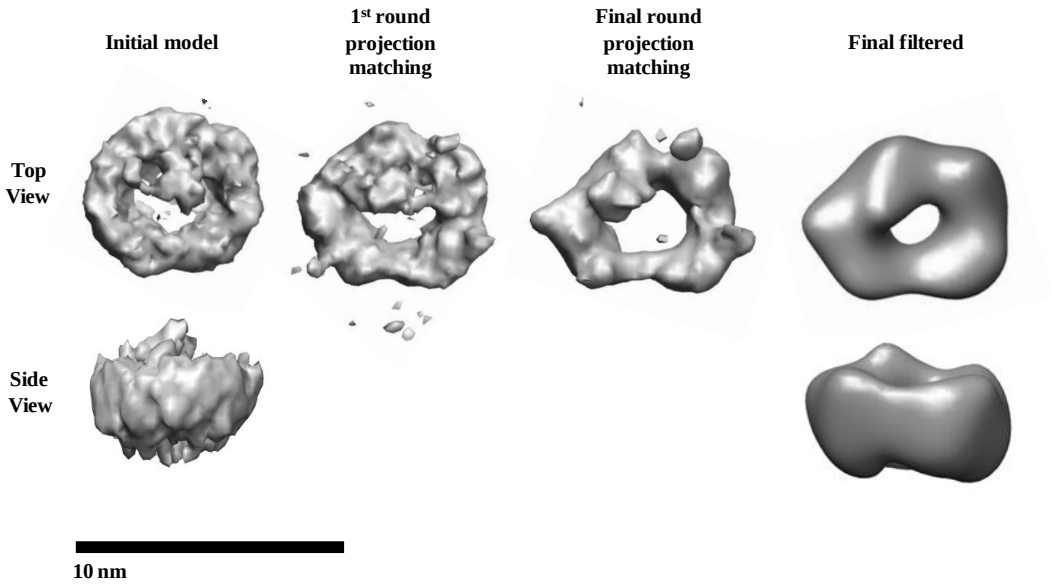
During this procedure any anomalous tilted particles, that showed little to no similarity with any of the projections, were discarded. A resolution check using the conservative 0.5 Fourier shell correlation (FSC) was performed after each round of refinement, and iterations were automatically stopped when no improvement was detected. The final reconstructions were generated using the more accurate simultaneous iterative reconstruction technique (SIRT). A final low-pass Butterworth filter was then applied to the final models around the nominal resolution.

The results of this refinement procedure are shown for each class in Figure 3.7.2. In each case the initial model consists of a relatively featureless ring. Over the course of the projection matching the globular ring densities become more apparent. The graph in Figure 3.7.2 displays the average improvement in FSC over the refinement procedure with the resolution improving from ~ 60 Å to ~ 30 Å after projection matching.

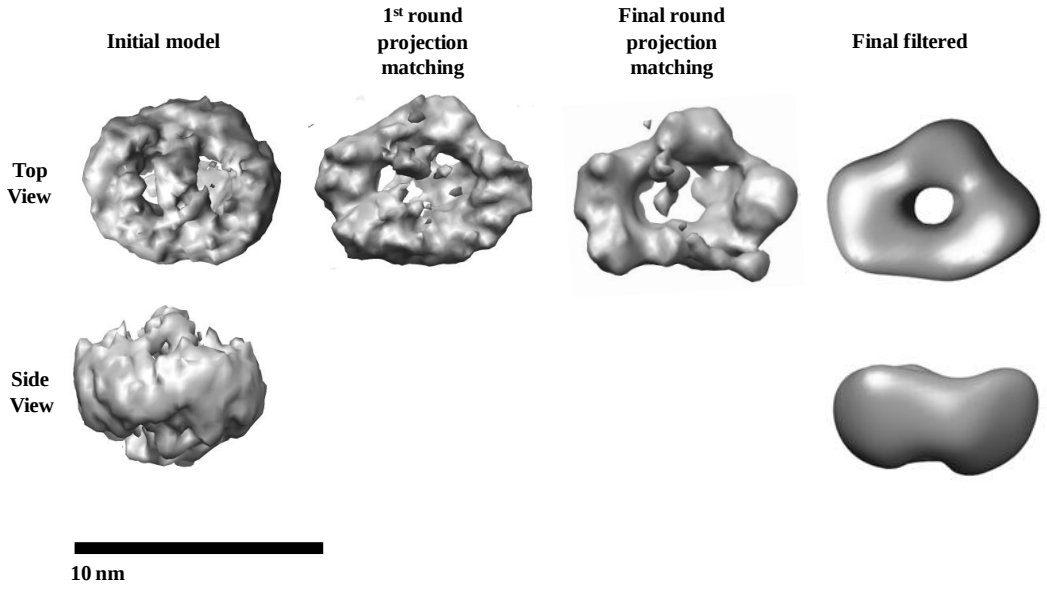
**Class 1**



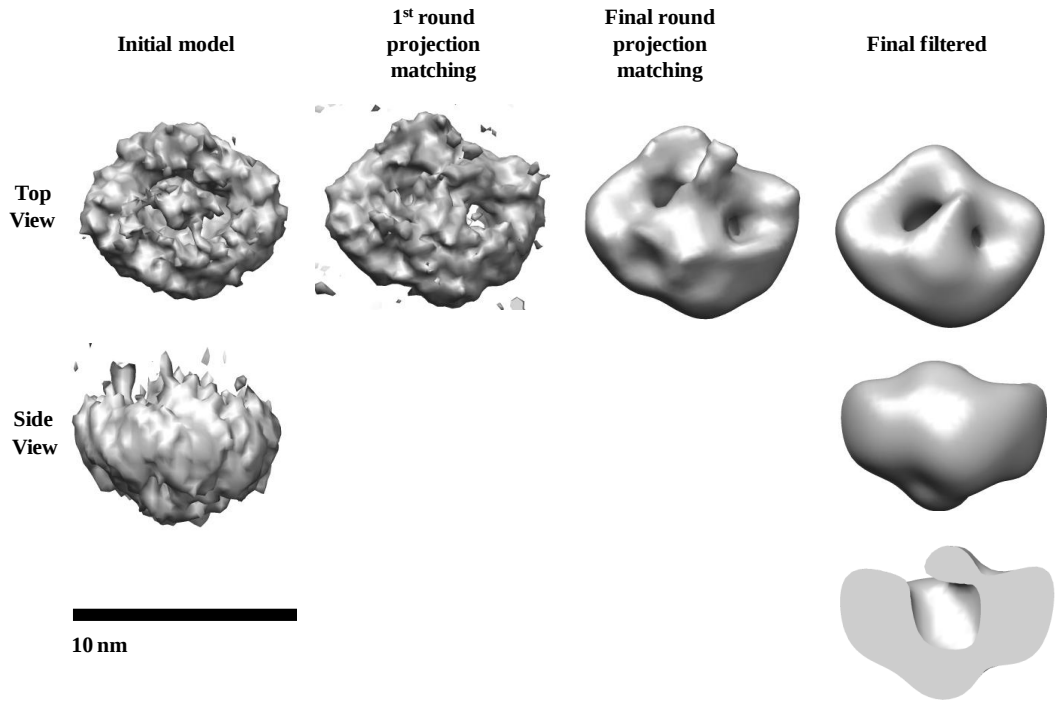

**Class 2**



**Class 3**




**Class 4**



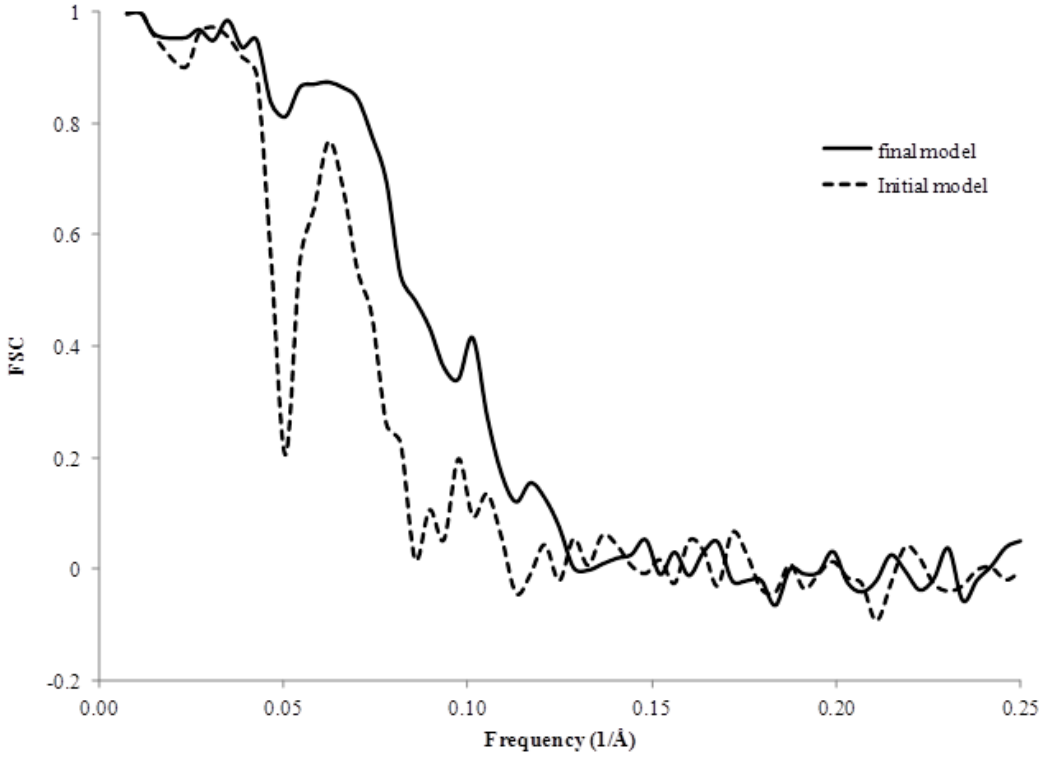



**Figure 3.7.2. Refinement of 3D models by projection matching**
For each class a 3D electron density map is shown following the stages of the projection matching refinement procedure. The final filtered model with all extraneous density removed is shown on the far right contoured to ~ 4 σ (standard deviations above the mean density). All contouring levels between the classes are relative. The side views represent a 90° rotation about the x axis. 3D models were displayed using UCSF Chimera (Pettersen *et al.*, 2004). The graph displays the average improvement in resolution over the refinement. The 0.5 Fourier shell correlation (FSC) was used for resolution estimates based on 2.62 Å/p.

The final filtered class averages are shown in Figure 3.7.3 along with the 2D class averages. The models show a clear resemblance to the class averages as validated by their 2D projections. The 3D density maps show small asymmetrical rings consisting of approximately 5 globular densities measuring ~ 2.5 nm across. No variation in the ring thickness is seen between the classes. Only in class 4 could a significant occlusion of the central channel be detected. In the other classes any potential 'lid' density did not align during the refinement and was subsequently removed during filtering. The central density in class 4 is more prominent on one face than the other as can be seen in the cross-section. The height of the ring measures ~ 4 nm for the three smaller classes, increasing to 5.5 nm for class 4. Using the enclosed volume at a contour level of ~ 4 σ the complexes are estimated to range from 70 kDa – 120 kDa. This is based on a helical protein arrangement of 0.844 Da/Å$^3$.

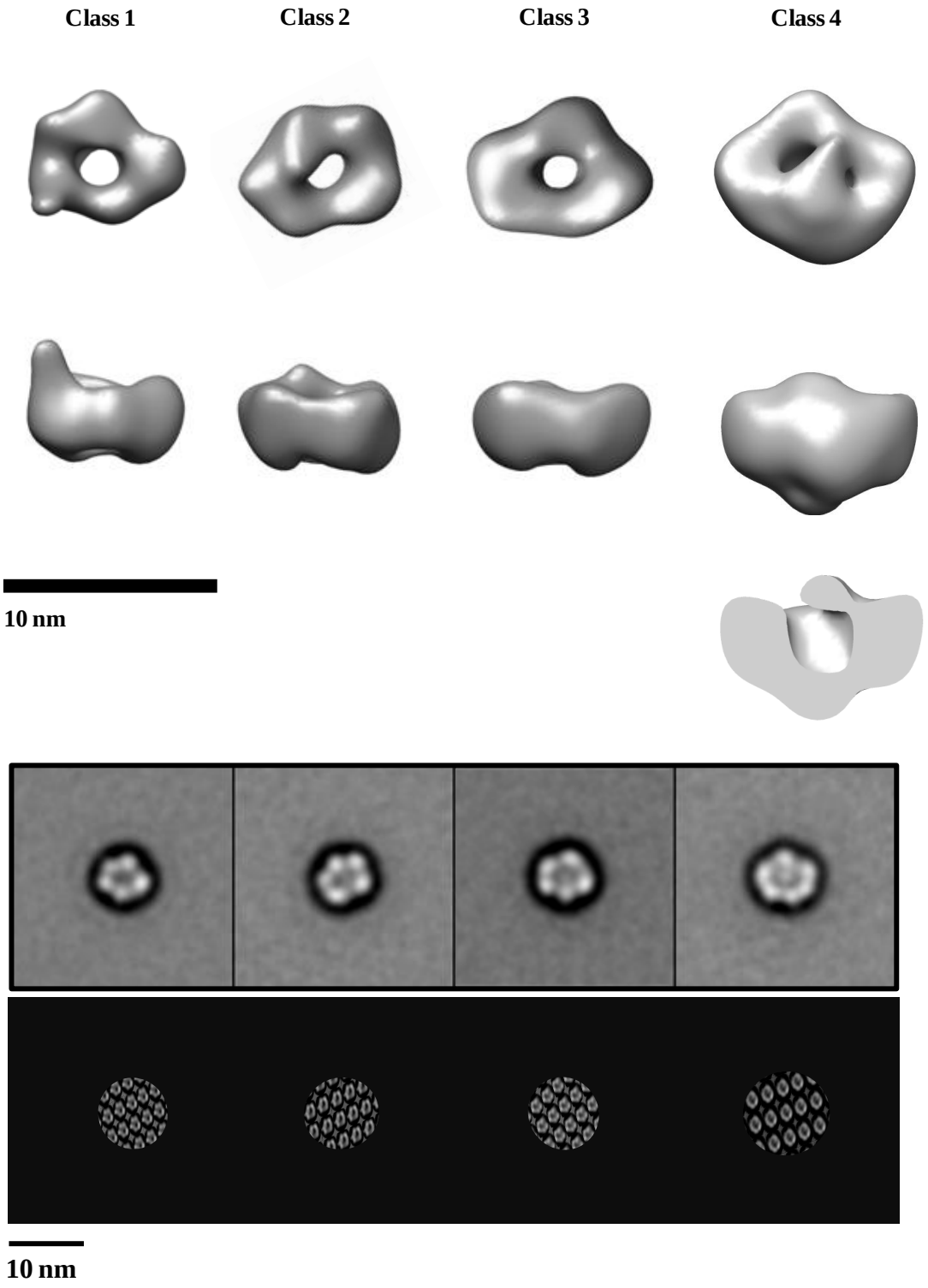



**Figure 3.7.3. Final 3D models of TatAd-*his***
The final reconstruction for size classes 1-4 are shown from left to right. The density maps are filtered to 30 Å and contoured to ~ 4 σ (standard deviations above the mean density). Side views are shown with the potential cytoplasmic face pointing down. Class 4 is also shown as a cross-section to reveal the internal cavity. The middle panel displays the aligned 2D class averages for reference. The lower panel shows 2D projections of the final models for comparison with the class averages.

Figure 3.7.4 shows the results of manually fitting the published solution state NMR structure of TatAd (Hu *et al.*, 2010) into the electron density map of class 4. The arrangement of the subunits follows the established theory with the 3 nm long transmembrane helix (TMH) fitting into the ring density and the 4 nm long amphipathic helix (APH) forming the 'lid' density. Previous structural work on TatAd monomers using both circular dichroism and solid-state NMR have indicated that the APH is oriented parallel to the lipid bilayer with a potential tilt (Lange *et al.*, 2007; Müller, Sonja D. *et al.*, 2007; Walther *et al.*, 2010) and the more recent solution-state NMR study has revealed a high degree of flexibility within the C-terminal region of the APH (Hu *et al.*, 2010). This may explain the difficulty in accurately resolving the 'lid' density across the size classes.

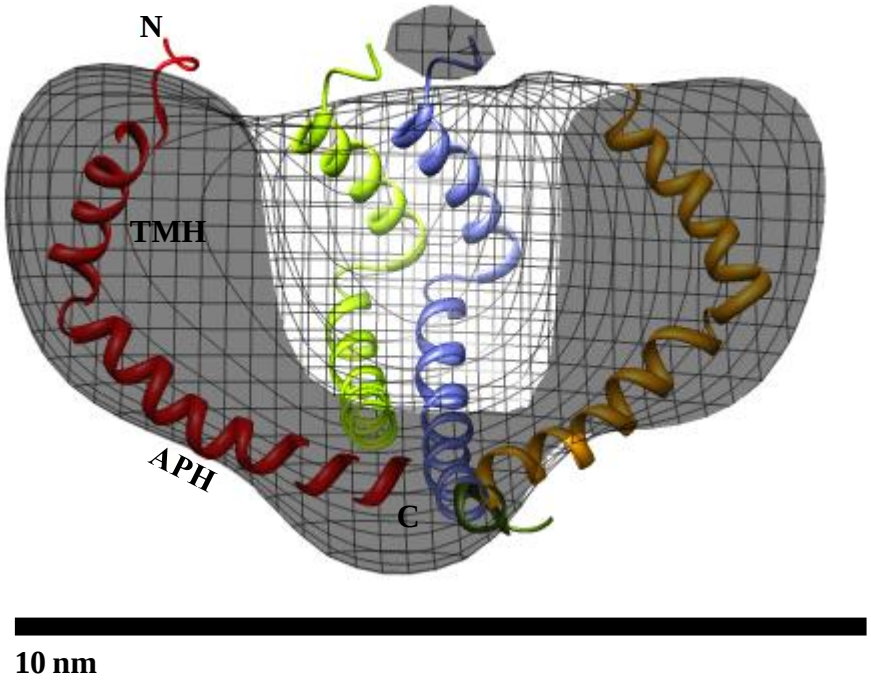



**Figure 3.7.4. Manual fitting of NMR solution structures of TatAd monomers**
A cross-section of the class 4 TatAd density map is shown. Copies of the NMR solution structure of TatAd (Hu *et al.*, 2010) were manually fitted into the ring using the relevant scaling options. The N and C termini of the TatAd monomer, as well as the position of the transmembrane helix (TMH) and amphipathic helix (APH) are indicated on the left.

During the refinement of the class 1 model extra density was seen to align producing extensions arising from the globular ring densities (see Figure 3.7.2 and 3.7.3). This alignment occurred on repeat of the procedure and could not be removed by tighter masking of the model or the projections used for refinement. On manual analysis of the class 1 tilted particle set, several examples of these extensions could be seen in the raw images, as shown in Figure 3.7.5. One, or in some cases two, short extensions measuring 1-2 nm in length, can be seen protruding from the ring density; resulting in a total length of ~ 4 nm over these extended regions, including the ring density. These extensions could not be detected in any of the other classes.

It is not clear what these extensions represent and they could potentially be formed by the binding of another protein. However, to date, and within this sample, no other TatAd binding proteins have been identified. As such it was presumed that this extra density was formed by the Tat protein. Due to the hydrophobic properties of the transmembrane helix of TatAd and the flexibility of the hinge region it is likely that, these extensions contain the amphipathic helix (APH). Based on the established conformation of the transmembrane helix within the lipid bilayer and ring structure of the TatAd complex, this suggests a rotation of the APH as shown in Figure 3.7.5.

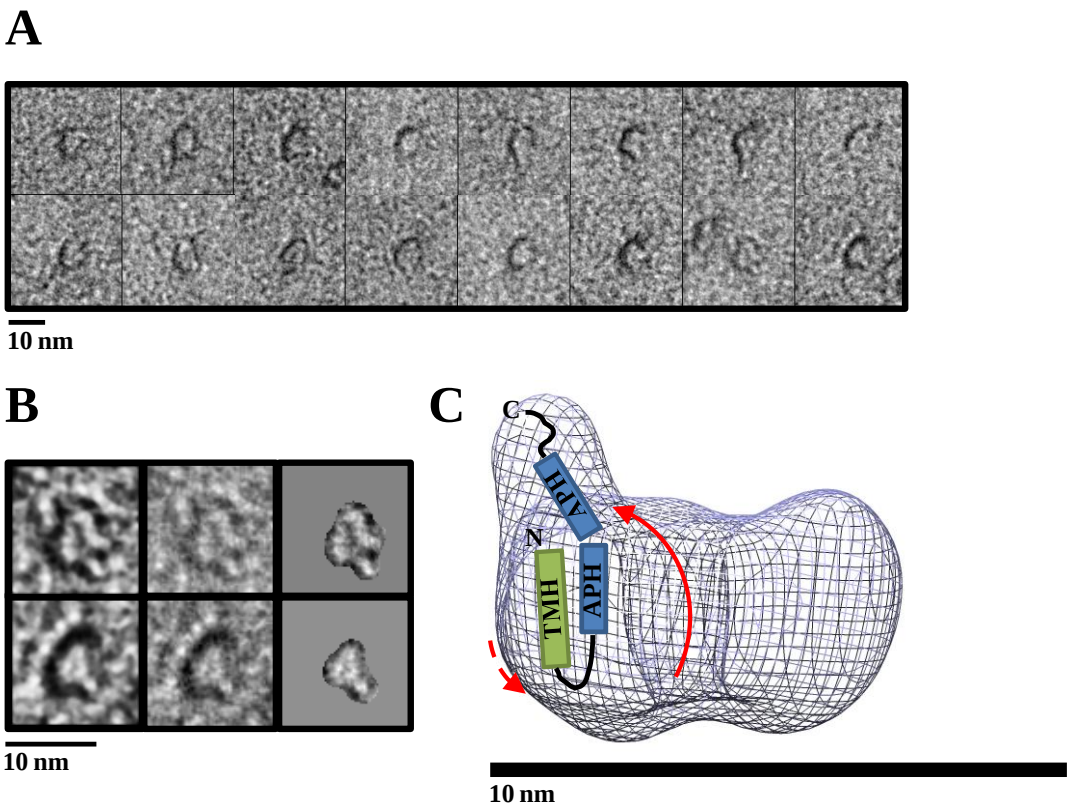



**Figure 3.7.5. Investigation of extensions from TatAd-*his* Class 1 ring structure**
*A.* The tilted particles contributing to the class 1 model were manually examined and several examples of potential extensions from the ring were observed. *B.* Close-up view of 2 of the particles from *A*, the top image potentially showing 2 extensions. From left to right: filtered image, raw image, masked image. *C.* Suggested conformational arrangement of TatAd subunit to occupy 'arm' density. The full arrow represents a hypothetical bending of the hinge region to move the position of the amphipathic helix (APH). The dashed arrow represents a shift in the position of the transmembrane helix (TMH) from its predicted tilted arrangement in the membrane to a straighter alignment. N and C represent the respective termini of the protein.

Such a conformational shift within the subunit has been proposed previously for *E. coli* TatA complexes (Gouffi *et al.*, 2004). In this case accessibility of the C-terminus to proteolysis both *in vitro* and *in vivo* was used as an indicator for dual topology. This change in conformation could potentially account for the reduced ring diameter in class 1 compared to the other classes. In class 4 the ring diameter is maximal and a clear 'lid' is present as shown in Figure 3.7.4, whereas in class 1 no lid density is observed but the extensions are. Based on these observations it is tempting to assign these as open and closed channel conformations. However, the reduction in ring and

channel diameter in the 'open' class 1 makes such an assignment unlikely as the complex is not large enough to accommodate fully folded Tat substrate.
In the absence of TatCd and substrate it is unclear how much functional significance can be imposed upon the observed structures. However the correlation between ring size and central or 'lid' density does suggest that a conformational change is evident. In the presence of TatCd, to form a structural scaffold, it may allow the 'open' state TatAd to form larger pores; whereas without such a scaffold the open state TatAd condenses to form a smaller and more energetically favourable, but not biologically relevant, complex. The flexibility of the APH and C-terminus of TatAd may explain the difficulty to resolving these extensions within the intermediate classes 2 and 3.

## 3.9. Discussion

Initially most structural studies of the bacterial Tat system have focused on the Gram-negative organism *E. coli*. More recently the minimal AC-type Tat systems of *B. subtilis* have gained interest due to their apparent structural homogeneity and the interesting comparison that can be drawn with the *E. coli* system

The structure of *B. subtilis* TatAd in a monomeric state has been solved by solution-state NMR (Hu *et al.*, 2010) and has been found to be in agreement with previous bioinformatic structural predications. However TatAd complexes have not previously been subjected to structural analysis.

TatAd complexes elute from gel filtration with a peak at 270 kDa but the more accurate size estimates of 3D models show complexes of 70-120 kDa. This demonstrates how much of an impact the detergent has on gel filtration and the inherit discrepancies in using the elution of soluble proteins to estimate those of membrane proteins in detergent.

The data shown here indicates that detergent concentration can have a very large effect on both the size and shape of particles observed by EM, and that a lot of caution must be taken when washing away excess detergent before imaging. Reduction of the detergent concentration within the membrane localised TatAd sample to below the CMC results in the formation of diverse particle populations.

These varied particles appear to match well with those seen in a previously observed soluble TatAd population (Westermann *et al.*, 2006).

TatAd was found to form ring-shaped structures of approximately 7.5-9.0 nm in diameter, containing a central pore of 2.5-3.0 nm. This represents significantly smaller complexes and a reduced size range of complexes, than those identified previously for *E. coli* TatA (Gohlke *et al.*, 2005). TatA complexes were found in a large size range of 9-13 nm with a corresponding variance in pore diameter of 3.7 nm. This led to the proposal that such a range in pore size would allow the complex to translocate the size range of known Tat substrates in a folded state. Importantly the largest size class of particles (with the largest pore) would be just wide enough to accommodate TorA a large 90 kDa Tat substrate.

The complexes observed for TatAd do not form such large complexes and no evidence of a pore of 7 nm was seen. The average particle diameter is smaller, at 8 nm, than the smallest observed class of TatA complexes. Both the biochemical and structural data presented here agree with previous reports that TatAd forms discrete complexes that vary considerably from the highly heterogeneous TatA complexes of *E. coli*. The biological relevance of this size range has been disputed by the observation that TatAd can substitute for *E. coli* TatA (Barnett *et al.*, 2008) to allow for translocation of the TorA substrate and can therefore interact successfully with the native *E. coli* TatBC complex.

The slight variation in TatAd complex size was seen to correlate with the presence or absence of proposed lid density on the cytoplasmic face of the complex. The larger complexes displayed clear lid density while the smallest complexes showed no lid density but the presence of extensions from the globular ring densities. It is therefore suggested that this represents a conformational change of the subunits within the complex. This conclusion is based on various studies into the orientation of TatA and TatAd within the membrane where the ability of the amphipathic helix to transiently flip into the membrane has been proposed.

# *Chapter 4*

# *Single-particle EM analysis of TatE complexes*

## 4.1. Introduction

As discussed previously, the majority of work on Tat systems has focused on the work-horse model bacterium *Escherichia coli*. Within this system three integral membrane proteins, TatA, TatB and TatC, comprise the minimal apparatus required for translocation; therefore, the focus of research has mostly remained fixed on these components.

However, *E. coli* also contains a monocistronic TatA paralog, TatE, that is thought to have arisen due to a gene duplication (Yen *et al.*, 2002). The *tatE* gene is found only in enterobacteria (Yen *et al.*, 2002), with the exception of the Gram-positive bacteria *Corynebacterium glutamicum* and *Corynebacterium efficiens* (Ikeda and Nakagawa, 2003). *E. coli* TatE is a 67-amino acid single transmembrane-span protein with 53% sequence identity to TatA (Sargent *et al.*, 1998).

TatE has largely been overlooked within the literature as the protein is normally present at much lower abundance than that of TatA (Jack, R. L. *et al.*, 2001). This low abundance, coupled with the apparently redundant role of TatE in complementing for a *ΔTatA* mutation (Sargent *et al.*, 1999), has left TatE relegated as an evolutionary hangover. However, it has been noted that a defective TatE strongly retards the translocation of some Tat substrates suggesting the potential for a divergent function from TatA (Sargent *et al.*, 1998).

However, the ability of TatE to complement for TatA, along with the high sequence similarity between the proteins suggests that there should be a conservation of any function-critical structural features. An in-depth investigation of TatE can therefore be used to further inform our understanding of the bacterial Tat system.

Interestingly, evidence of non-redundant roles for TatE also exists. In *E. coli*, *tatE* is induced during biofilm formation (Heikkila *et al.*, 2001), and in *Pseudomonas stutzeri* the expression of *tatE* is associated with the process of denitrification (Beloin *et al.*, 2004). These studies suggest TatE may play an important role under specific growth conditions, or that it may constitute a substrate specific adaptation.

In this chapter the results of the first structural investigation into *E. coli* TatE complexes is presented. This project was undertaken in collaboration with Jacopo Baglieri, whom performed all purification and biochemical assessment prior to the

EM imaging and single-particle analysis. *E. coli* TatE complexes were purified following expression in the presence of native levels of TatB and TatC.

Initially the results of the TatE complex purification are summarised. All figures of TatE gels (Figure 4.2.1 and Figure 4.2.2) are the work of Jacopo Baglieri and are presented here with his permission to provide a reference for the EM imaging.

This is followed by the results of experiments to optimise on-grid imaging conditions.

Next the results of a single-particle EM analysis of TatE complexes are presented; this study includes the separation and classification of differently shaped and sized TatE complexes. These data indicate that the majority of TatE complexes form very small complexes with little size variation, results that match well with the biochemical data. Potential multi-ringed structures and side views are also separated and analysed.

Together these results indicate that *E. coli* TatE complexes are far smaller and more homogeneous than TatA complexes, with a central pore not large enough to translocate large Tat substrates like TorA. These observations suggest that the inferred relevance of the TatA complex size range may be less functionally significant than previously thought. This has important implications for the mechanism of Tat translocation as a whole.

The TatE complex was very challenging to work with due to its small size and low abundance compared to other purified Tat complexes. All stages of the imaging process required multiple optimisations and overall sample purity and particle heterogeneity were a constant issue. To this end, various dilutions, gridding approaches and washing steps were experimented with. As with TatAd (chapter 3), the protein:detergent ratio was observed to be vital for controlling aggregation and particle morphology. The work presented in this chapter has been published in part (Baglieri *et al.*, 2012) and the publication is provided in Chapter 7.

## 4.2. Purification of TatE complexes

Initially purification of TatE complexes was performed using $C_{12}E_9$ detergent. Under these conditions the sample was not isolated to a high purity, although the resulting EM imaging revealed a population of small and discrete particles. However, these appeared to lack definition of internal features and were heterogeneous in size and shape (~ 8-20 nm); as such they were not well suited for single particle work. Analysis of the original Streptactin™ elution fraction revealed larger globular particles varying in size up to ~ 50 nm diameter as well as even larger amorphous particles of over 300 nm. These likely represent excess detergent, lipid or other contaminating protein. This gave a clear indication that the gel filtration step of the purification procedure was essential before EM experiments were performed. Attempts to improve the quality of the imaging data by dilution, washing steps or changing stains were unsuccessful

To try and address these problems a repeat purification was performed, using the smaller micelle sized DDM detergent instead. In this case similar particles were observed although imaging conditions were much improved; the background staining is lower and the particles have improved boundary definition and contrast. Particles still appear heterogeneous in size and shape with the majority of particles appearing small and round (10 nm or less down to ~ 5 nm) with evidence of a central cleft or cavity, along with rod shaped particles (~5 nm by 11-13 nm) and larger globular particles up to 20 nm in diameter. Multi-ringed structures/multimers of smaller ringed structures could also be seen. Under these conditions the sample still appeared too heterogeneous for a meaningful single particle analysis.

A 3$^{rd}$ repeat of the purification, again in DDM, resulted in a very pure sample as shown in Figure 4.2.1. The gel filtration elution profile of this sample shows that TatE elutes as a small, discrete set of complexes with an average size of ~ 130 kDa, based on soluble protein standards. However, as stated previously, this estimation is rough at best and is markedly impacted by the detergent micelle. The silver-stained PAGE of the peak elution fraction (GF23) indicates that contaminating proteins were present at very low levels.

A more accurate size assessment by BN-PAGE was also conducted successfully, suggesting that only 2-3 different sized complexes of 50-110 kDa were present, as

shown in Figure 4.2.2. This represents a drastically reduced size range of complexes from those published previously for TatA, where complexes of less than 50 kDa to over 500 kDa have been observed (Alami *et al.*, 2003; Bolhuis *et al.*, 2001; Oates *et al.*, 2005).

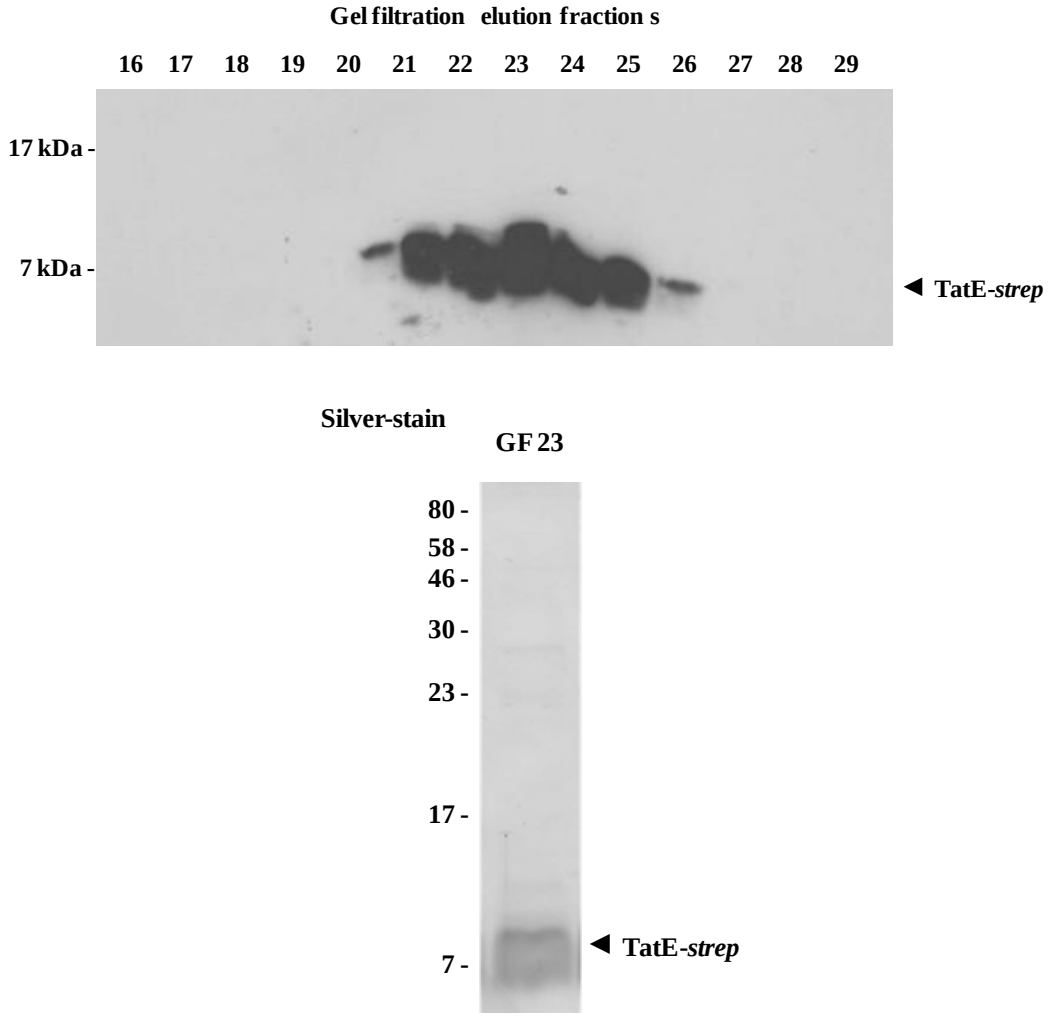



**Figure 4.2.1. Purification of TatE-*strep* in DDM by gel filtration**
Membranes were isolated from *E. coli ΔtatAE* cells expressing TatE with a C-terminal *Strep*-II™ tag. Affinity purified TatE containing fractions were applied to a Superdex™ 200 10/300 GL gel filtration column as described previously (see TatAd chapter). The peak gel filtration elution fractions were analysed by immunoblotting against the *Strep*-II™ tag (top panel), and the purity of fraction 23 was assessed by silver-stain (lower panel). The position of TatE is indicated on the right of the figure and molecular weight markers are shown on the left.

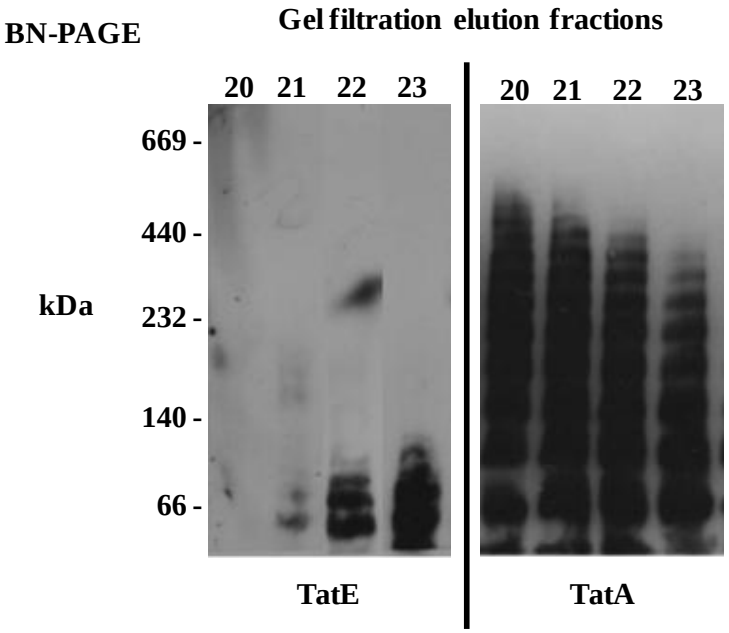



**Figure 4.2.2 BN-PAGE analysis of TatE vs. TatA complexes**
Gel filtration elution fractions of TatE and TatA were subjected to blue-native (BN) PAGE. Unlike the continuous ladder of TatA bands, TatE forms three low molecular weight complexes around ~ 66 kDa. The protein complexes were detected by immunoblotting with antibodies against the C-terminal *Strep*-II™ tag on TatE and TatA. Molecular weight markers are shown on the left.

## 4.3 Optimisation of EM grid conditions for TatE-*strep* complex visualisation

Gel filtration fraction 23 was used for all subsequent EM image analysis work. Grids were produced using the same touching drop method (see methods, Chapter 2) used for the TatAd sample (Chapter 3). An image of the undiluted sample is shown in Figure 4.3.1. Small, round particles are present, and these appear discrete and well defined. However, distinct sub-populations are still evident, including short rods and some larger particles. Under these conditions the grid appears too densely populated for single particle analysis and required further sample dilution.

Figure 4.3.2 shows the results of a 1:2 dilution of TatE in detergent-free buffer. Small, round particles are present in a very large continuous spectrum of sizes, from ~ 5 nm to over 40 nm in diameter. Such a size range is not shown in either the gel filtration data or BN-PAGE results (see Figure 4.2.1 and Figure 4.2.2) indicating that this size variance has been introduced during gridding and is potentially artefactual. Such a variation of particle dimensions in relation to changes in detergent concentration was also observed for the TatAd results shown previously (Chapter 3).

This grid represents the best imaging conditions achieved with this sample in terms of contrast, but unfortunately due to the observed change in particle morphology it was inappropriate to use for downstream processing.

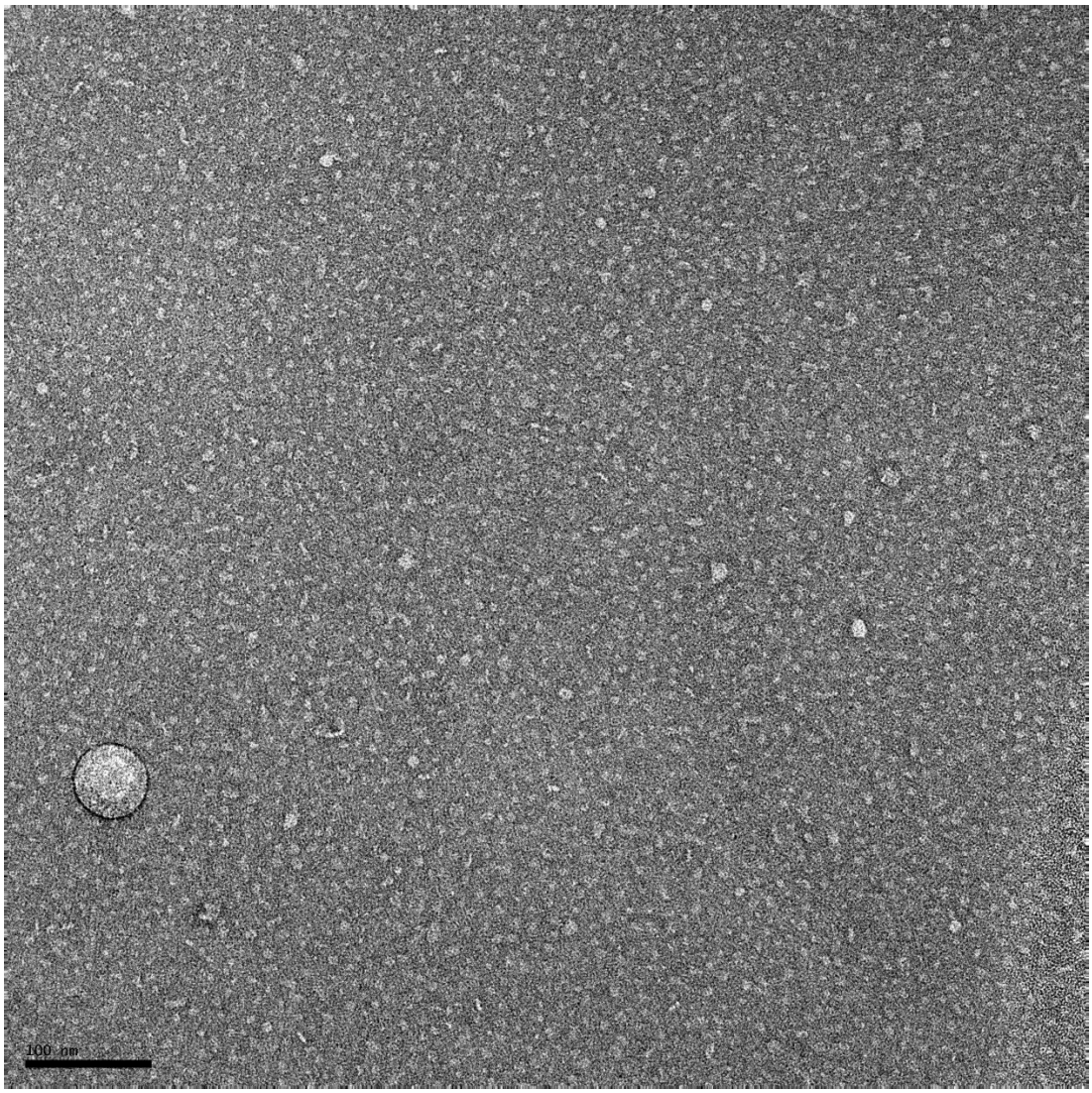



**Figure 4.3.1. Negative-stain micrograph of TatE-*Strep* undiluted**
The micrograph was taken at ~ 71,500x magnification under ~ 1.5 µm defocus. The grid was stained with 2% uranyl acetate. The grid is densely packed with small, round particles, along with short rod-shaped structures and some larger particles. Scale bar = 100nm

When TatE is diluted 1:2 with GF buffer + 0.02% DDM the distinct size variation is drastically reduced, as shown in Figure 4.3.3., supporting the conclusion that the previous dilution was affecting the particle appearance. The majority of particles appear small and round with a diameter of < 10 nm and a clear central pool of stain. However, short rods and multi-ringed particles are still present. The large, amorphous structures seen in Figure 4.3.2 are no longer present.

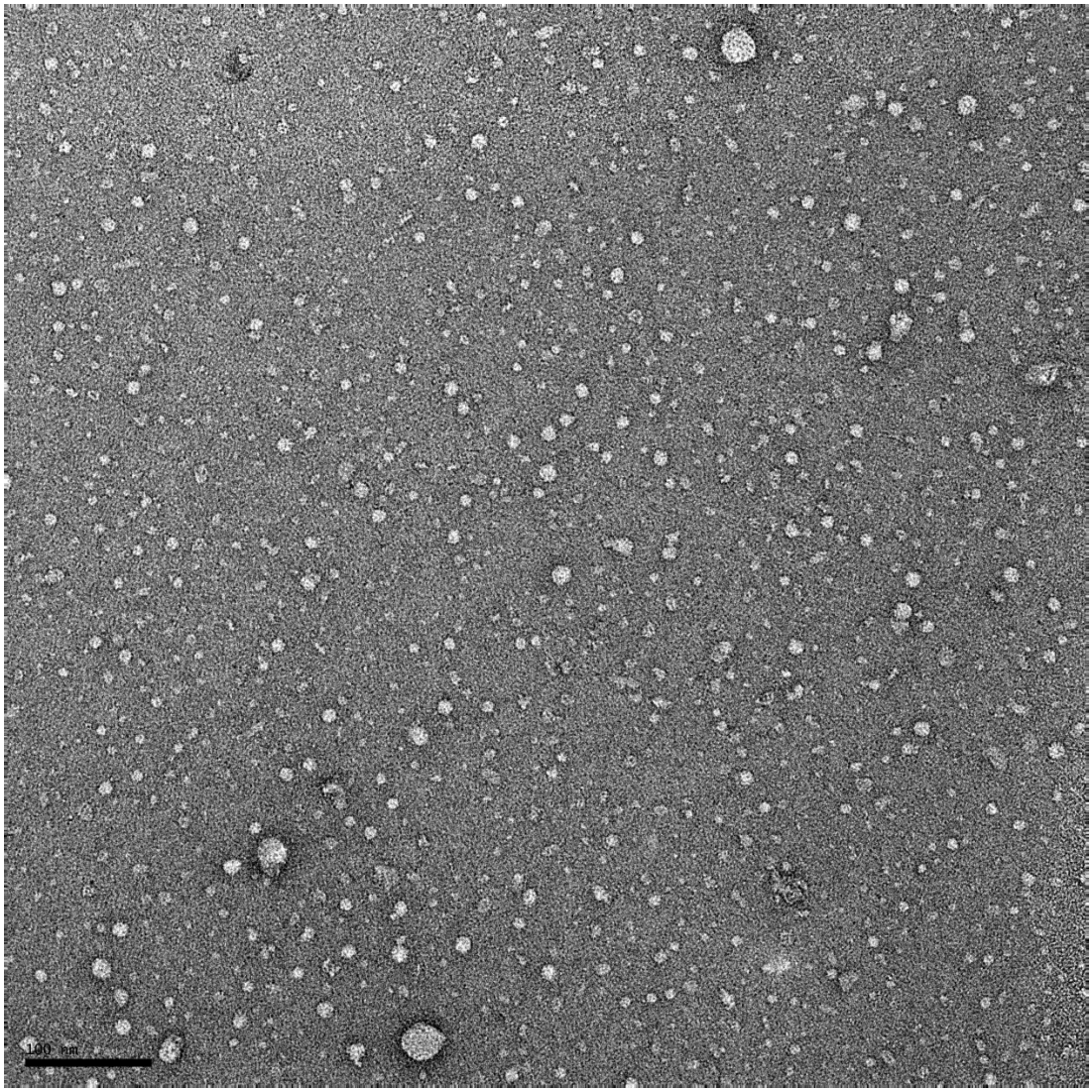



**Figure 4.3.2. Micrograph of TatE-*Strep* 1:2 dilution in detergent-free GF buffer**
The micrograph was taken at ~ 71,500x magnification under ~ 1.5 µm defocus. The grid was stained with 2% uranyl acetate. Scale bar = 100nm

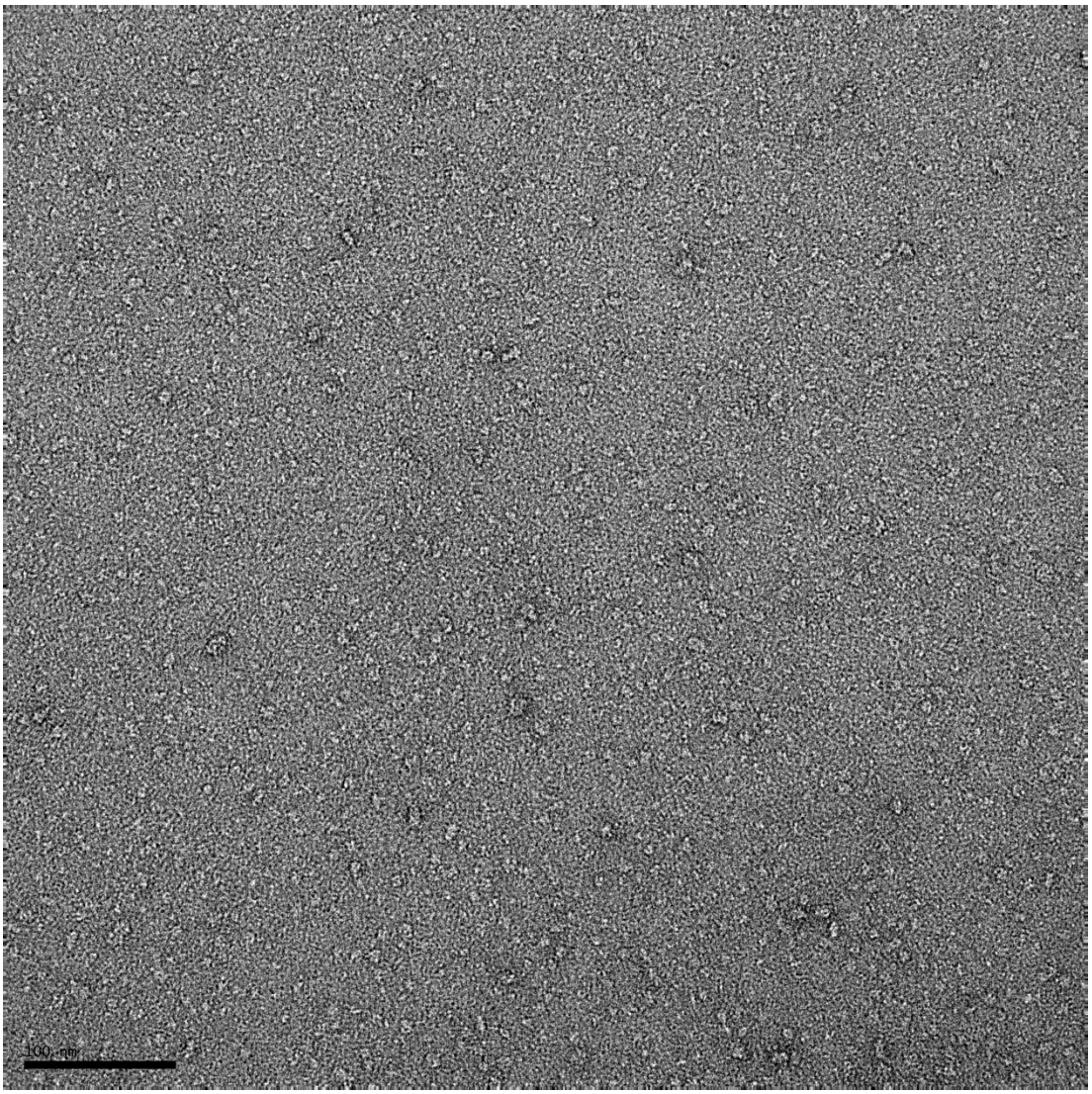


**Figure 4.3.3. Micrograph of TatE-*Strep* 1:2 dilution in GF buffer + 0.02% DDM**
The micrograph was taken at 85,878x magnification under ~ 1.5 µm defocus. The grid was stained with 2% uranyl acetate. Scale bar = 100nm

The imaging conditions shown in Figure 4.3.3 are not optimal for particle-picking, as the contrast is fairly low and the particle density is high (compare with Figure 4.3.2). However, repeated attempts to dilute the TatE sample further or introduce additional washing steps led to a plethora of different artefacts and adverse imaging conditions, examples of which are shown in Figures 4.3.4 and 4.3.5.

The results of diluting TatE 1:3 and washing three times with $dH_2O$ are shown in Figure 4.3.4. In this case a large size range of particles, similar to those shown in

Figure 4.3.2, are produced indicating that the protein may be aggregating due to the reduced detergent to protein ratio.

Additional grids of the 1:2 dilution with GF buffer + 0.02% DDM (Figure 4.3.3) were also produced, an example of which is shown in Figure 4.3.5. In this case the background staining appears uneven producing a mottled effect on the grid. These conditions lead to a lot of localised variance in contrast levels and are therefore unsuitable for particle picking as this will introduce artefactual variance during latter classification steps. On the other hand, these additional gridding attempts do indicate that the majority population of small, round particles is stable and reproducible (compare Figure 4.3.3 and Figure 4.3.5).

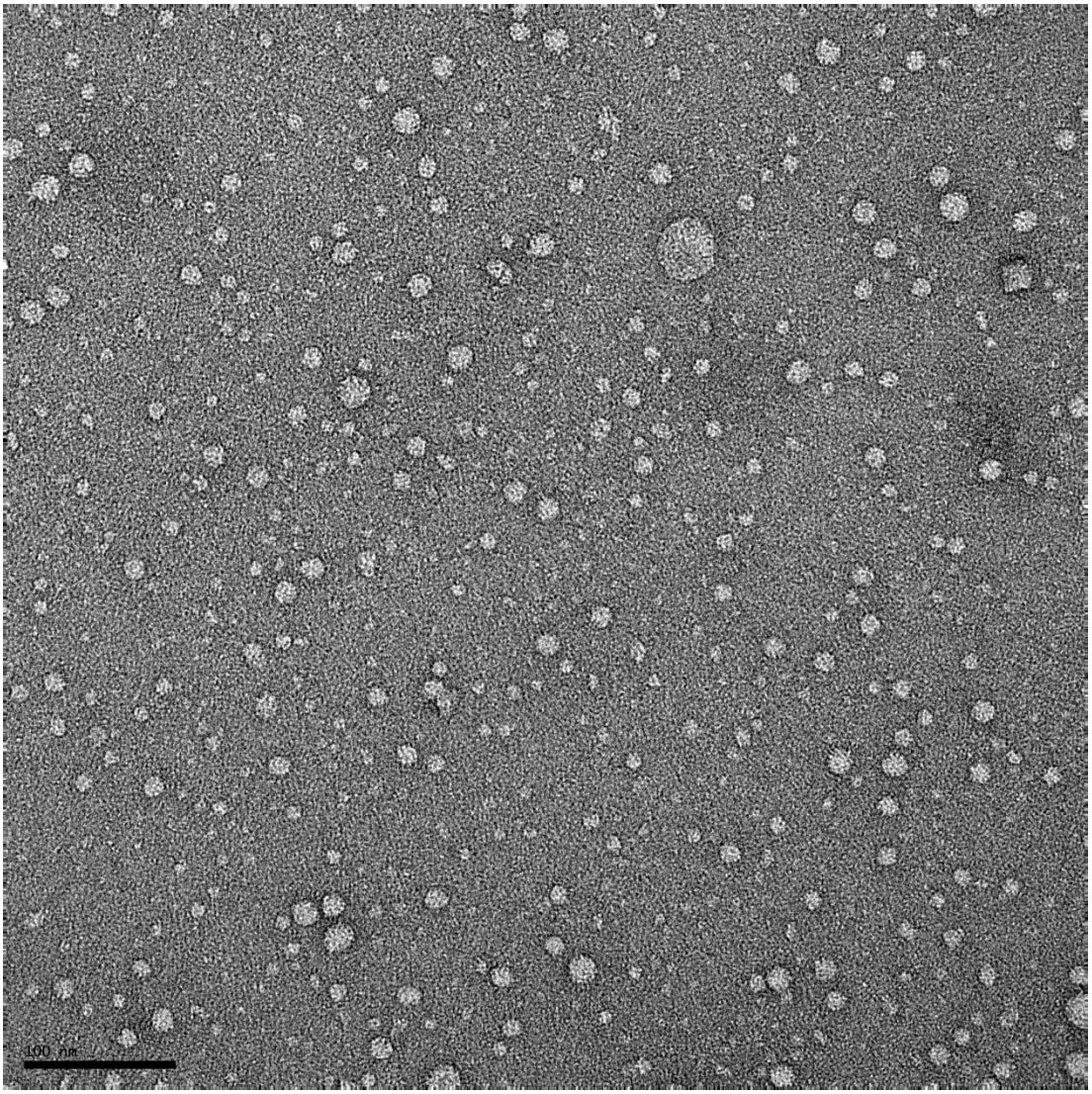



**Figure 4.3.4. Micrograph of 1:3 TatE-*Strep* + 3x $dH_2O$ wash**
The micrograph was taken at 85,878x magnification under ~ 1.5 µm defocus. The grid was stained with 2% uranyl acetate. Scale bar = 100nm

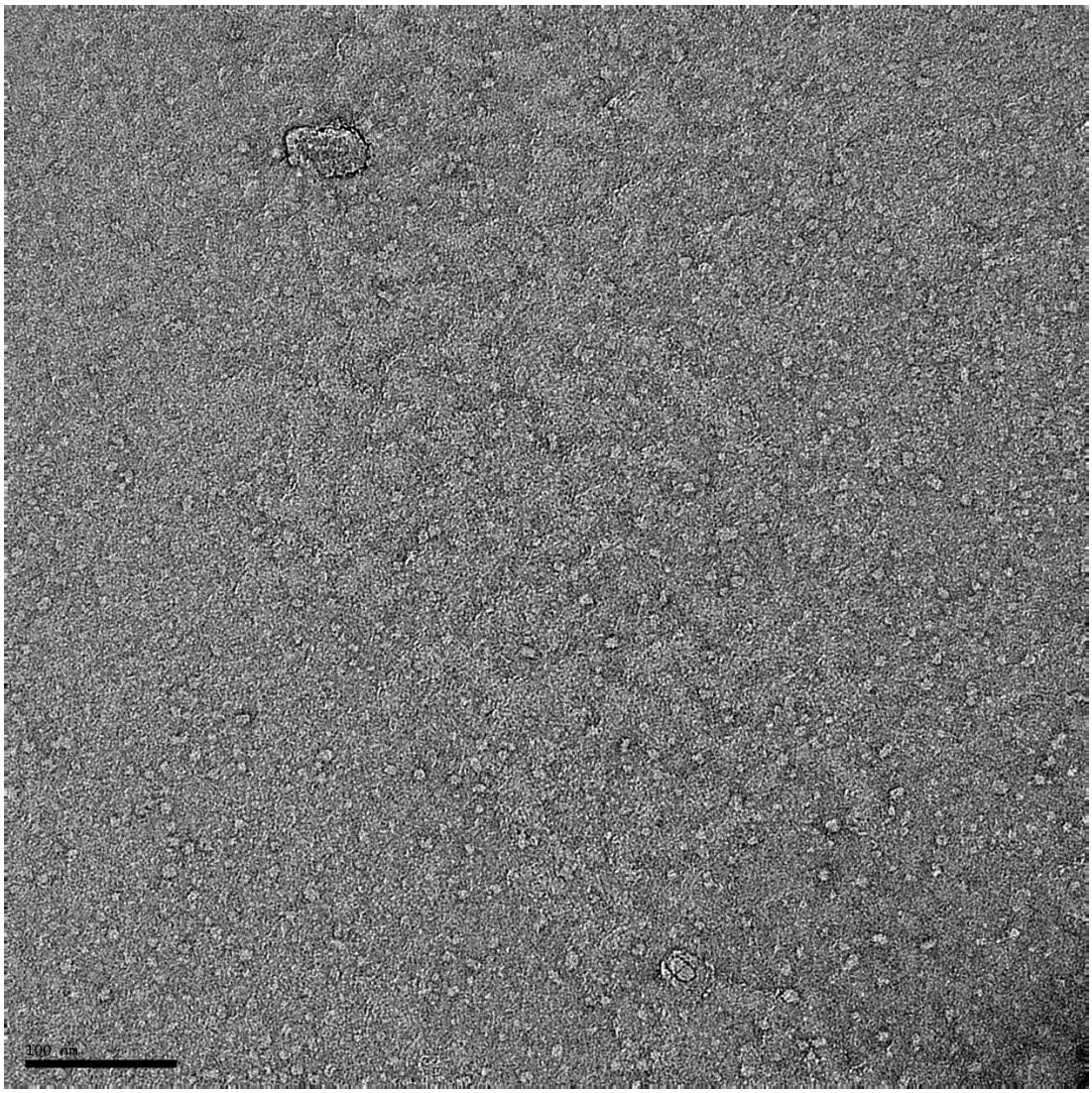



**Figure 4.3.5. Micrograph of TatE-*Strep* 1:2 dilution additional attempt**
The micrograph was taken at 85,878x magnification under ~ 1.5 µm defocus. The grid was stained with 2% uranyl acetate. Scale bar = 100nm

## 4.4. Single-particle analysis of TatE-*strep* complexes

The distribution and reproducibility of the TatE particles as shown in Figure 4.3.3 was deemed appropriate to continue forward into single-particle analysis. Therefore, a tilt-series of new images were acquired, covering a representative proportion of the grid, to provide sufficient particles for classification and the relevant views for downstream 3D reconstruction.

### 4.4.1. Particle picking and crude classification

Tilt-pairs were collected from the 1:2 dilution of TatE in GF buffer + 0.02% DDM at ~ 86,000x magnification, with a tilt angle of 45°, as shown in Figure 4.4.1. This tilt angle was used to avoid the contrast-fringe artefact seen in the 50° tilted TatAd data-set (see Chapter 3). The average defocus across the selected tilt pairs was 1.8 µm, a higher defocus was used as the particles are very small and do not present much contrast at lower defocus levels making them very difficult to distinguish from the carbon background.

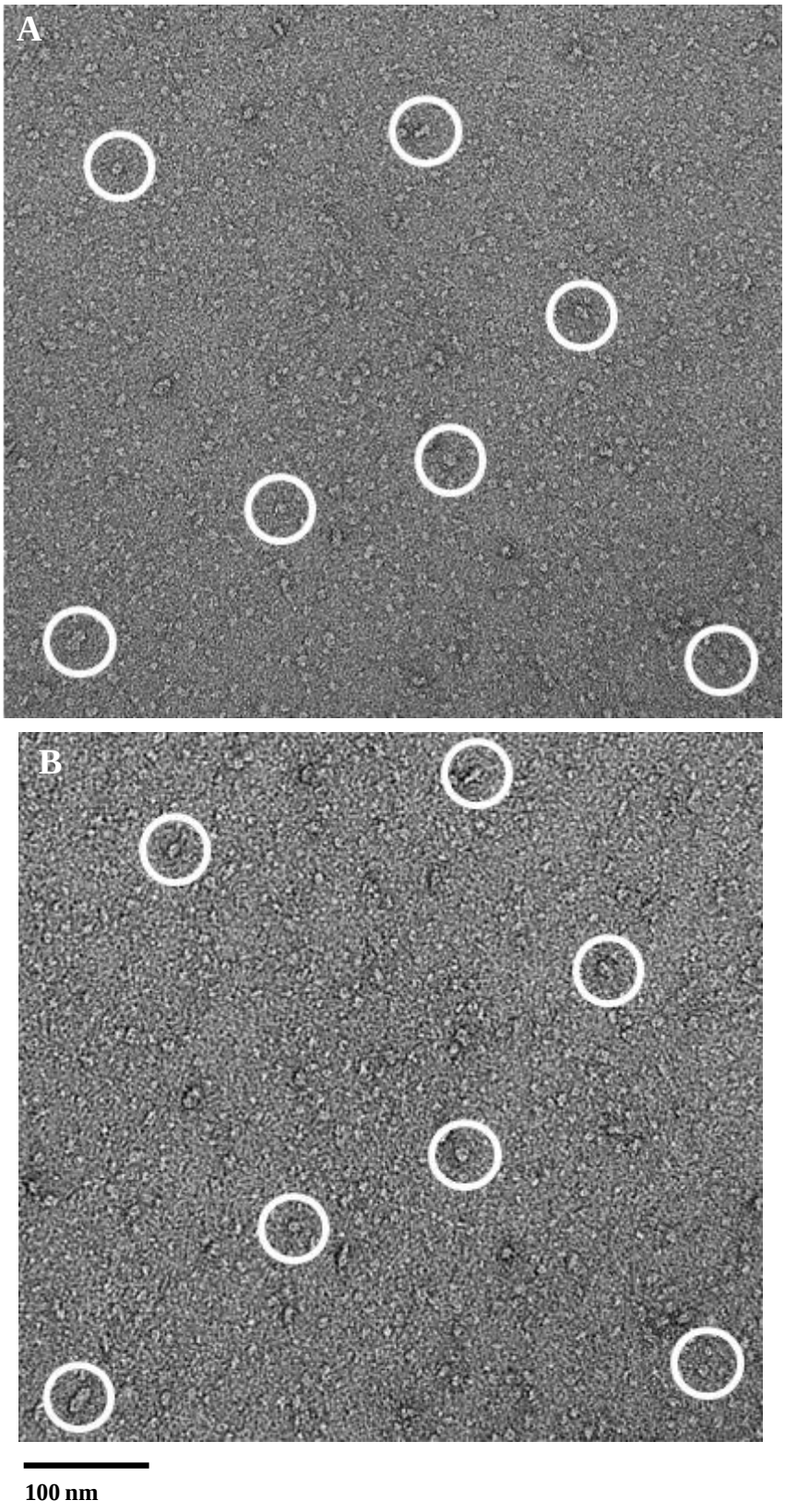



**Figure 4.4.1. Micrographs of TatE-*strep* 45° tilt-pair**
Sections of two micrographs covering the same area of the grid are shown. *A.* microscope stage untilted (0°). *B.* Micrograph stage tilted to 45°. Corresponding particles are circled across both images. Micrographs were taken at 85878x magnification under ~1.8 µm defocus. The grid was stained with 2% uranyl acetate.

The micrographs were assessed for image quality with reference to their 2D power spectra in the same way as the TatAd data-set (see Chapter 3). This resulted in the selection of 10 tilt-pairs of images. From these, 1012 initial particles were windowed out into 128x128 pixel boxes using the interactive particle picking tool in Jweb™.

Initial analysis of these particles revealed the majority of them to form ring-shaped structures with a maximal diameter of ~ 8 nm and a dark central stain pool indicating a potential pore or cavity. However, after multiple attempts it became clear that too much heterogeneity was present within the data-set to follow the eigenimage based size-separation method (White *et al.*, 2004) as used before for TatAd (see Chapter 3). Figure 4.4.2 shows the total centred average (T) and class averages (1-7) generated upon applying this classification method to the entire TatE particle set. Although size variance can be seen, most clearly by comparing 1 and 6, the averages appear blurred, off centre and badly defined. A number of factors contributed towards this issue: both different sized particles and potentially different views are present, the magnification and defocus are both increased compared to the TatAd sample and the particles appear on average to be smaller.

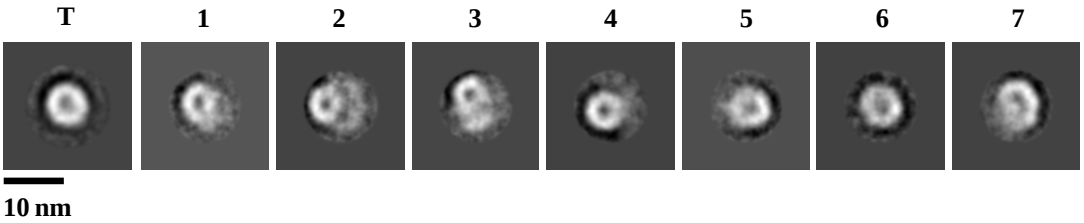



**Figure 4.4.2. Initial attempt at size-based classification on entire TatE data-set**
Centred average of total TatE particle set (T) and class averages (1-7) generated by HAC on eigenfactors showing a concentric circular pattern indicative of size variance.

To overcome these problems a step-wise method of cross-correlation was used to sequentially separate out different sub-populations prior to any size-based classification. First, only images showing a clear top down view of a ring-shaped structure were selected for analysis as these represent the vast majority of TatE complexes. In this way 190 particles were manually picked, examples of which are shown in Figure 4.4.3. These selected particles were then translationally aligned as

shown in Figure 4.4.4, and the centred average used as a cross-correlation reference for the rest of the data-set. Using this method for each of the 3 crude initial populations resulted in 259 ringed particles, 95 rod-shaped particles, and 132 multi-ringed structures, with the remaining 437 particles unclassified.

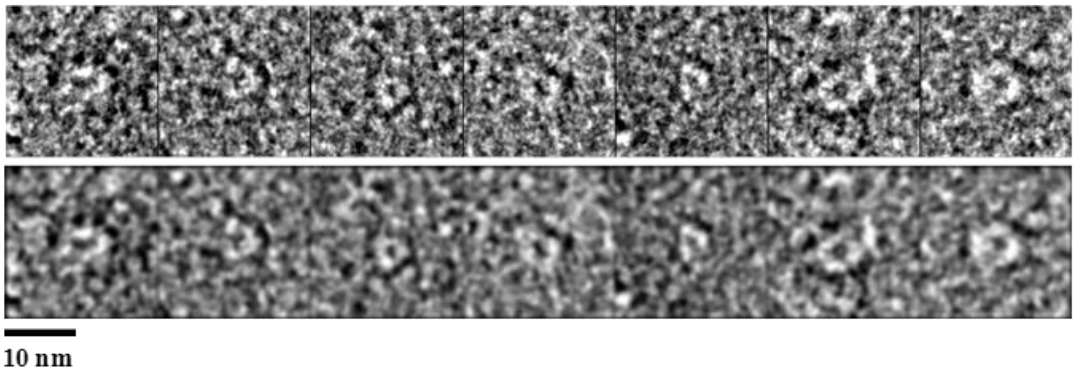



**Figure 4.4.3. Initially selected TatE-*strep* complexes with clear ringed shape**
Examples of manually selected TatE-*strep* complexes displaying a clear 'top down' view of a ring-shaped particle. The top panel shows raw, unprocessed images. The lower panel show the same particles processed with the *despeckle* tool of ImageJ (ver. 1.44p) to reduce the background noise.

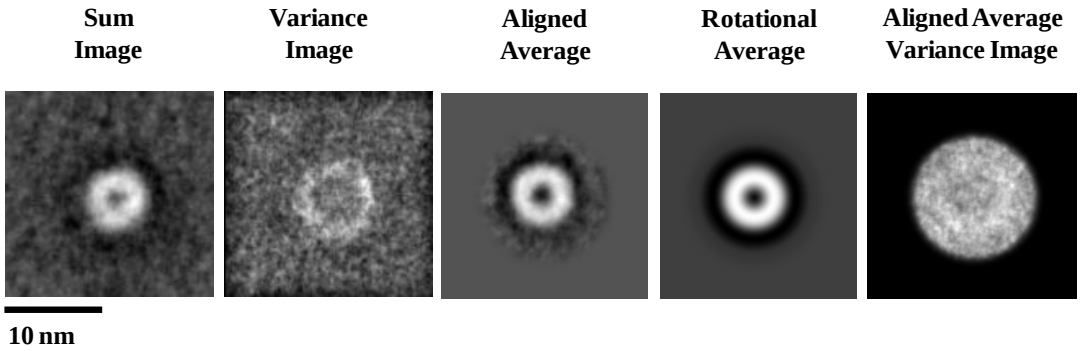
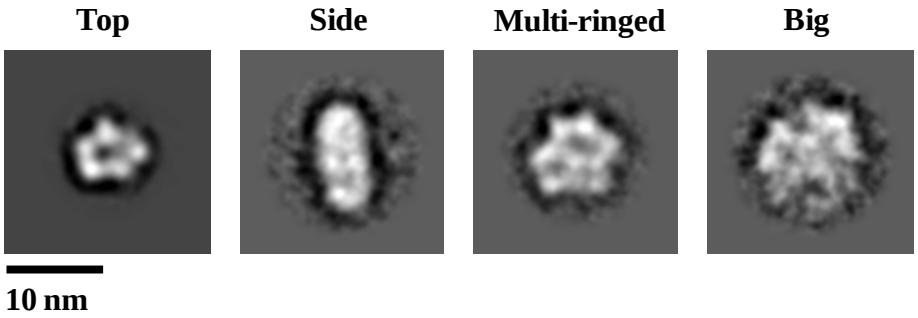



**Figure 4.4.4. Centring of manually selected TatE-*strep* complexes**
Averaged images of the 190 manually selected TatE particles displayed a clear ringed-structure. From left to right: Initial summed and variance images prior to alignment, translationally aligned average image and rotational average used as cross-correlation reference, variance image of aligned average showing reduced pixel variance.

### 4.4.2. Assignment of the unclassified particles

As the initial class assignment was very conservative, to simply separate the clearly disparate particles, a large number of particles remained unassigned. A reference-free translational alignment was applied to these filtered particles with no centring reference. To look for more ringed particles these centred particles were cross-correlated against the clear top-down average. The average cross-correlation value appeared low so a manual cut-off selection was set at 0.3 CCC (cross-correlation coefficient) resulting in only 70 selected particles. Reference-free alignment was run on the selected particles using the top-down view as a centring reference. Running the same procedure on the 367 rejected particles resulted in an average that resembled the multi-ringed set, indicating the population was still heterogeneous.

The rejected particles were manually assessed again and some much larger particles were observed and removed (49 particles in total); upon reference-free alignment the average of these shows a large ~ 11 nm wide, particle with no central pore. In contrast the remaining 317 particles produce a small, ~ 7 nm wide, average with a central stain pool but lacking a clear particle boundary. As such, these particles were cross-correlated against the original top-down average and the bottom 10% were removed, resulting in 291 additional small, round particles. After this process the crude separation of particles was distributed as shown in Figure 4.4.5.




| Class | Top | Side | Multi-ringed | Big |
|---|---|---|---|---|
| # Particles | 620 | 95 | 132 | 49 |

**Figure 4.4.5. Crude classification of TatE complexes based on gross-morphology**
The class averages shown have been rotationally and translationally aligned using a reference-free method. Class names were selected based on the suspected particle orientation or relevant structural feature. The numbers of particles per class are provided in the accompanying table.

### 4.4.3. Refinement of the crude classification

First an iterative centring method was applied to the corresponding original filtered particles, so as not to induce any bias from the prior alignments. A total of 3 iterations were performed with the size of the mask applied to the particles reduced after each round, as shown in Figure 4.4.5. The centred particles were then cross-correlated back to their centred average and the bottom 10% were removed, this edited set was then re-centred.

MSA (multi-variate statistical analysis) was then performed on the centred particles with the second eigenimage showing signs of size variation, as shown in Figure 4.4.6. HAC (Hierarchical ascendant classification) using this eigenfactor shows a very steep step after the first branching point and a great deal of divergence beyond the 4$^{th}$ branching point, indicating that only a small number of significant classes are present. Using a cut-off to produce 4 classes shows only 2 distinct sizes of average each with an apparent subclass. Therefore, a cut-off was selected generating only the 2 significant classes and the particles were re-centred accordingly. The bigger class measures ~ 7.7 nm across and the smaller class measures only ~ 6.0 nm. As shown in Table 4.4.3, the spread of particles between these classes is quite even suggesting that both represent significant and stable TatE complexes.

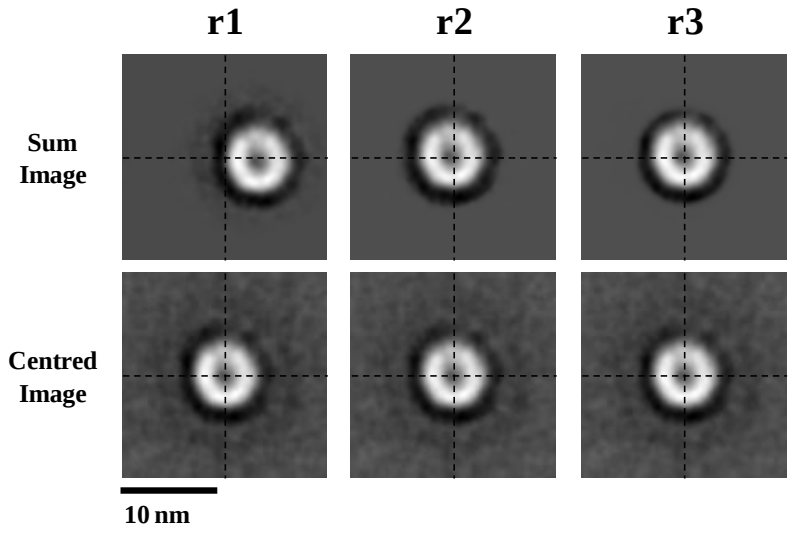



**Figure 4.4.5. Iterative centring of crude small, ringed class of TatE complexes**
The 620 selected TatE particles were centred via 3 iterative rounds (r1-3) of reference-free translational alignment. Resulting in a final centre of gravity of -0.02, -0.02 pixel units.

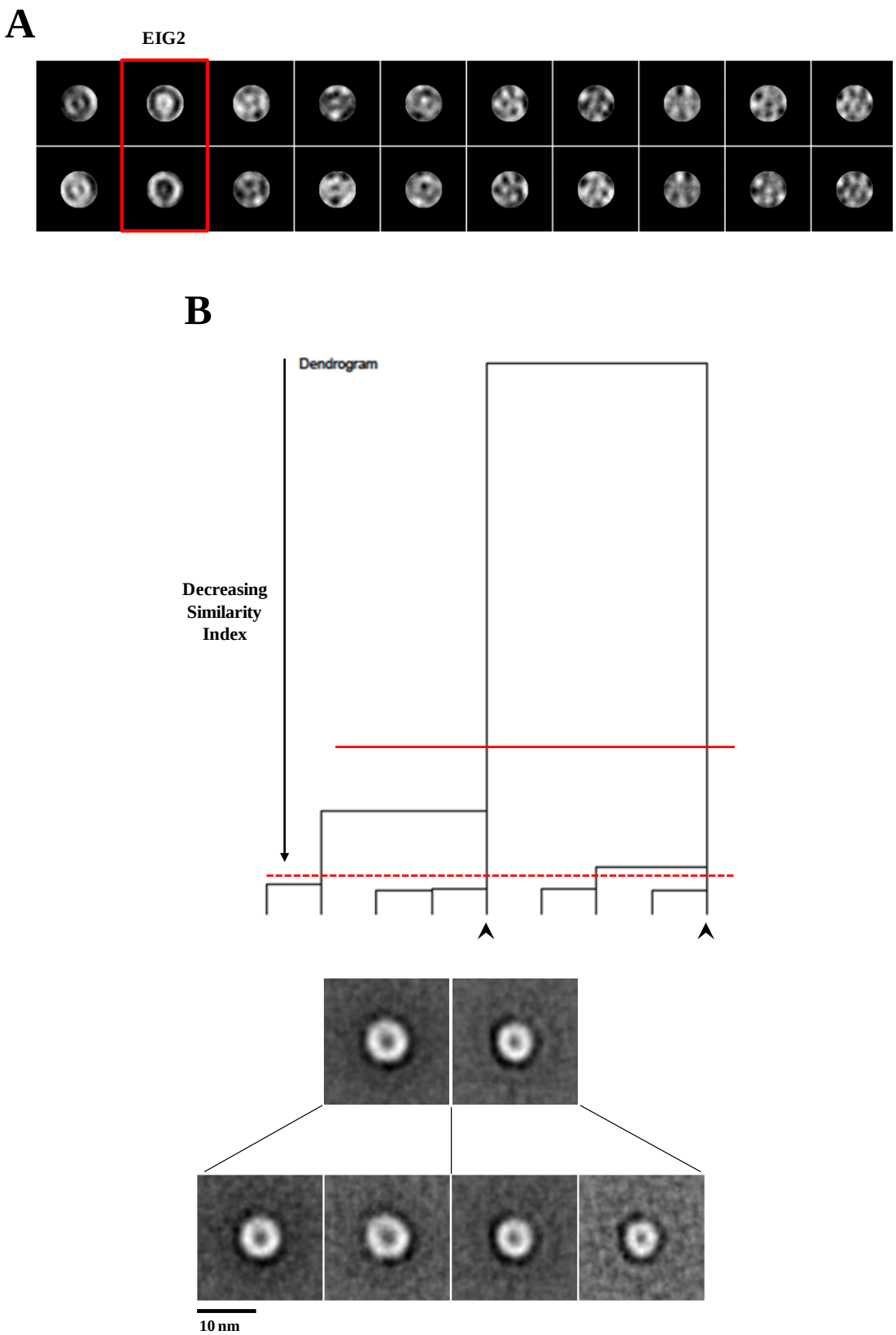



**Figure 4.4.6. Size classification of small, ringed TatE complexes**
*A.* MSA of the TatE particles revealed a concentric ring pattern indicative of size variation in eigenfactor 2 (EIG2) highlighted in red. *B.* HAC based on EIG2 revealed 4 potential size classes based on Ward's clustering method. A final cut-off value was selected generating 2 stable size classes.

| Class # | 1 | 2 |
|---|---|---|
| Diameter (nm) | 7.7 | 6.0 |
| # particles | 314 | 251 |
| # particles post CC edit | 288 | 226 |

**Table 4.4.3. Distribution of TatE particles between single-ringed size classes**
The size of each TatE class average and the number of constituent particles is shown. The size classes were edited by cross-correlation (CC) to the relevant class average with the lower 10% of correlating particles removed resulting in the reduced class populations as shown in # particles post CC edit.

The particles in each class were again cross-correlated back to their averages and the bottom 10% removed, as indicated in Table 4.4.1, before reference-free translational and rotational alignment. The tilted particles were centred using the same custom masking method described for the TatAd data-set (see chapter 4). The final 2D class averages and corresponding centred tilted particle sum images are shown in Figure 4.4.7. The size variation between the class averages can be clearly seen and is reciprocated in the tilted sum images. The ringed particles consist of several globular densities with ~ 4 visible in the smaller class (Class 1) and ~ 5 in the larger class (Class 2).

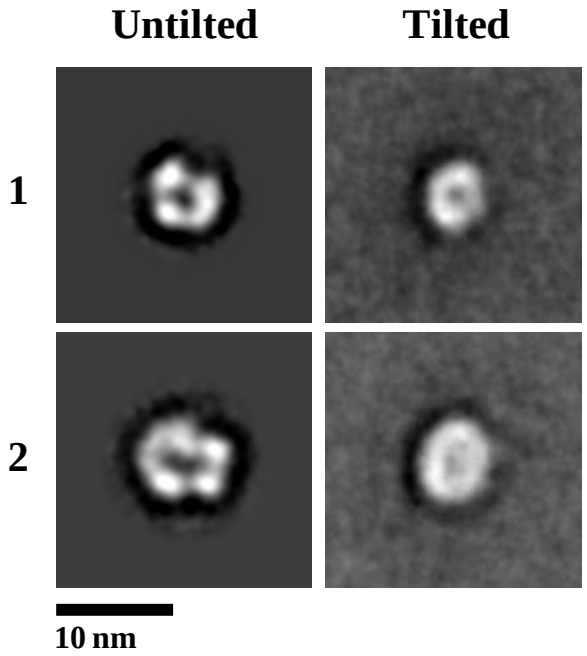



**Figure 4.4.7. Final 2D class averages of Single-ringed TatE-*strep* complexes**
The rotationally and translationally aligned class averages for the 2 size classes of small ringed complexes are shown. The smaller class 1 (226 particles) measures ~ 6 nm in diameter. The larger class 2 (288 particles) measures ~ 7.5 nm. Sum images of the centred tilted particles within each class are shown on the right.

### 4.4.4. Multi-ringed complexes of TatE-*strep*

The dimensions of the longer rod shaped particles suggest that they do not represent alternate views of the single ringed particles. Based on the longest dimension these rods correspond quite well with the average of the multi-ringed particles as shown in Figure 4.4.8. Based on these observations it seems likely that these rods represent side views of the larger structures. These larger multimers are more likely to be affected by the flattening induced during negative-stain grid preparation and would therefore adopt a wider array of overall shapes and conformations. Along with the increased size of the particles, these factors would make this sub-population less inclined to adopt a preferred orientation on the grid. In some cases it was possible to see the multi-ringed structures from a side-view, depending on the orientation of the particle relative to the tilt axis. As shown in Figure 4.4.9, a comparison of tilted and untilted particles shows the potential angular relationship between these sub-populations. This clustering of smaller ringed particles into larger structures with 2 or more defined rings is a property observed across all of the Tat complexes shown in this investigation (see Chapters 3 and 5)

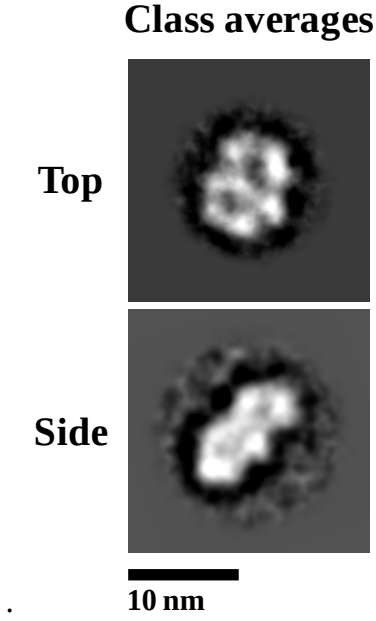



**Figure 4.4.8. 2D class averages of Multi-ringed TatE-*strep* complexes**
Rotationally and translationally aligned class averages for the 2 other significant particle classes are shown. These have been assigned as top down (134 particles) and potential side views (95 particles) of the same multi-ringed complex. Two distinct pores can be seen in the top view and these are occluded in the side view. The top view measures 6 nm wide reduced to 5 nm in the side view. In both cases the longest dimension measures 10.5 nm.

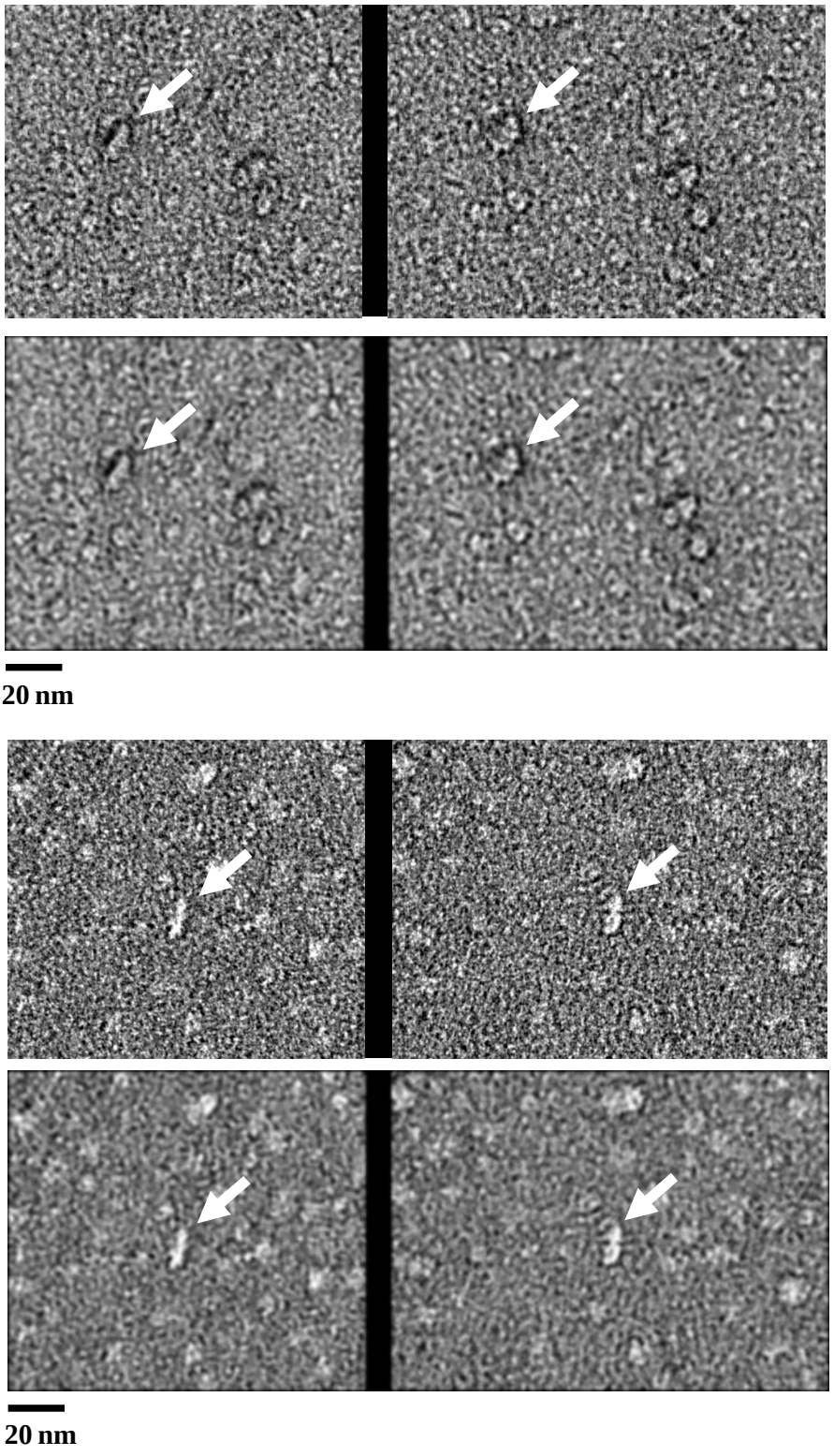



**Figure 4.4.9. Orientational relationship between larger TatE-*strep* complexes**
Two sections from different tilt-pairs of micrographs are shown, the images on the left are tilted to 45° and the right-hand images are untilted. In each case the top panel is unprocessed and the lower panel has been filtered to improve the S/N ratio. White arrows indicate the same particles tilted and untilted. In the untilted images the indicated particles are composed of 2 ringed structures, when tilted they resemble long rods of the same length.

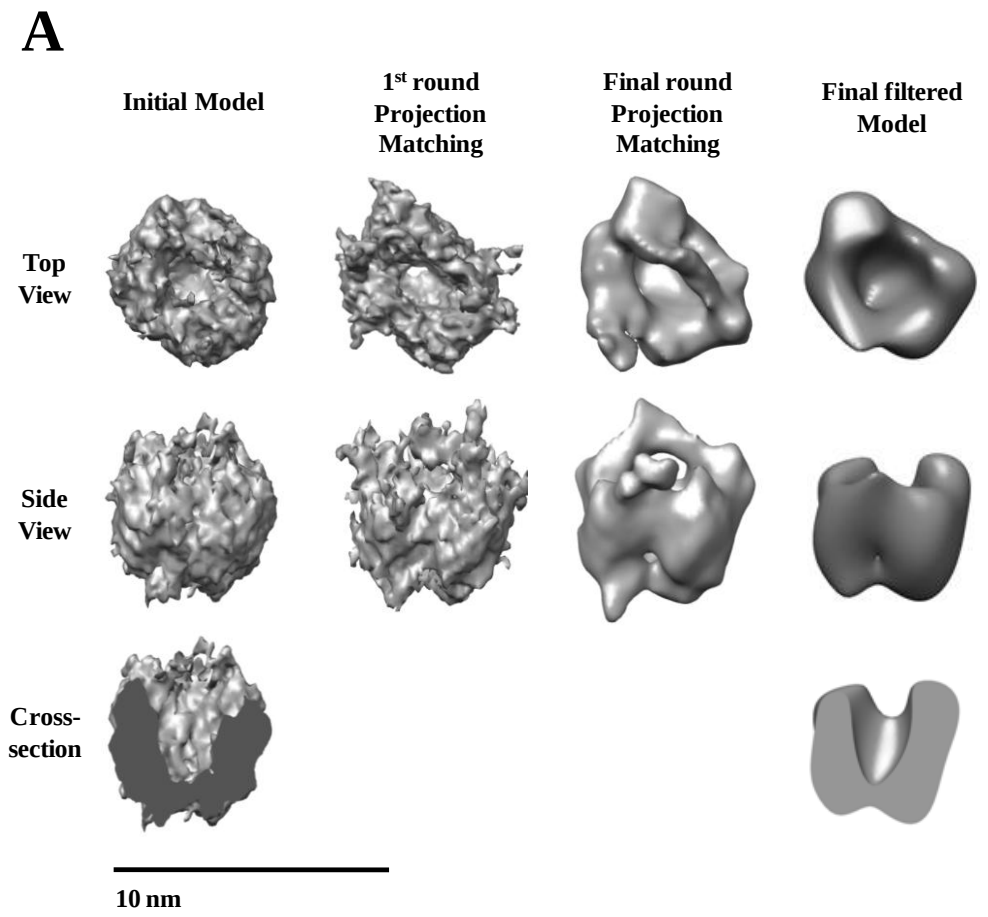
## 4.5. 3D reconstructions of TatE complexes

### 4.5.1. Two size classes of small, ringed TatE complexes

3D reconstructions of the TatE complexes were generated and refined using the same methods described for TatAd (chapter 3). Figure 4.5.1 shows several stages in the projection matching refinement procedure for both the small class (A) and the large class (B). As can be seen the initial models appear more rotationally averaged with a rounder more featureless appearance. Over the course of the projection matching the models become more structurally defined and the globular densities forming the ring become apparent, matching well with those seen in the 2D averages (Figure 4.4.7). Importantly the presence of the central cavity occluded on one face can be seen in the initial models and is conserved throughout the refinement.

The final 3D density maps (C) clearly show the small variation in overall size between the 2 classes of TatE complexes, less than 2 nm. The models are shown contoured to ~ 4.8 σ (SD of mean density) and filtered to 29 Å. The total height of both the complexes is ~ 5 nm, sufficient to span the plasma membrane. Interestingly, the 'pore' size for either complex appears approximately the same at 2.5 nm. This is far smaller than the ~ 7 nm wide pore recorded for the largest size class of TatA complex (Gohlke *et al.*, 2005); a size that was deemed sufficient to accommodate the large *E. coli* tat substrate TorA. Overall these data indicate that the TatE complexes would not be large enough to accommodate TorA via their central cavity. The small central cavity and dramatically reduced size range of TatE complexes casts doubt on the proposed 'size-fitting pore' model of Tat mediated translocation

The molecular weights of the complexes were estimated at 68 kDa and 90 kDa respectively, based on the enclosed density contoured at the ~ 4.8 level σ, using an alpha-helical protein packing of 0.844 Da/Å$^3$. These size estimates fit very well with the small, discrete bands observed by BN-PAGE (Figure 4.2.2) and together provide good evidence that these models are a reliable estimation of the purified TatE complexes.


A
Initial Model
1st round Projection Matching
Final round Projection Matching
Final filtered Model
Top View
Side View
Cross-section
10 nm
B
Initial Model
1st round Projection Matching
Final round Projection Matching
Final filtered Model
Top View
Side View
Cross-section
10 nm

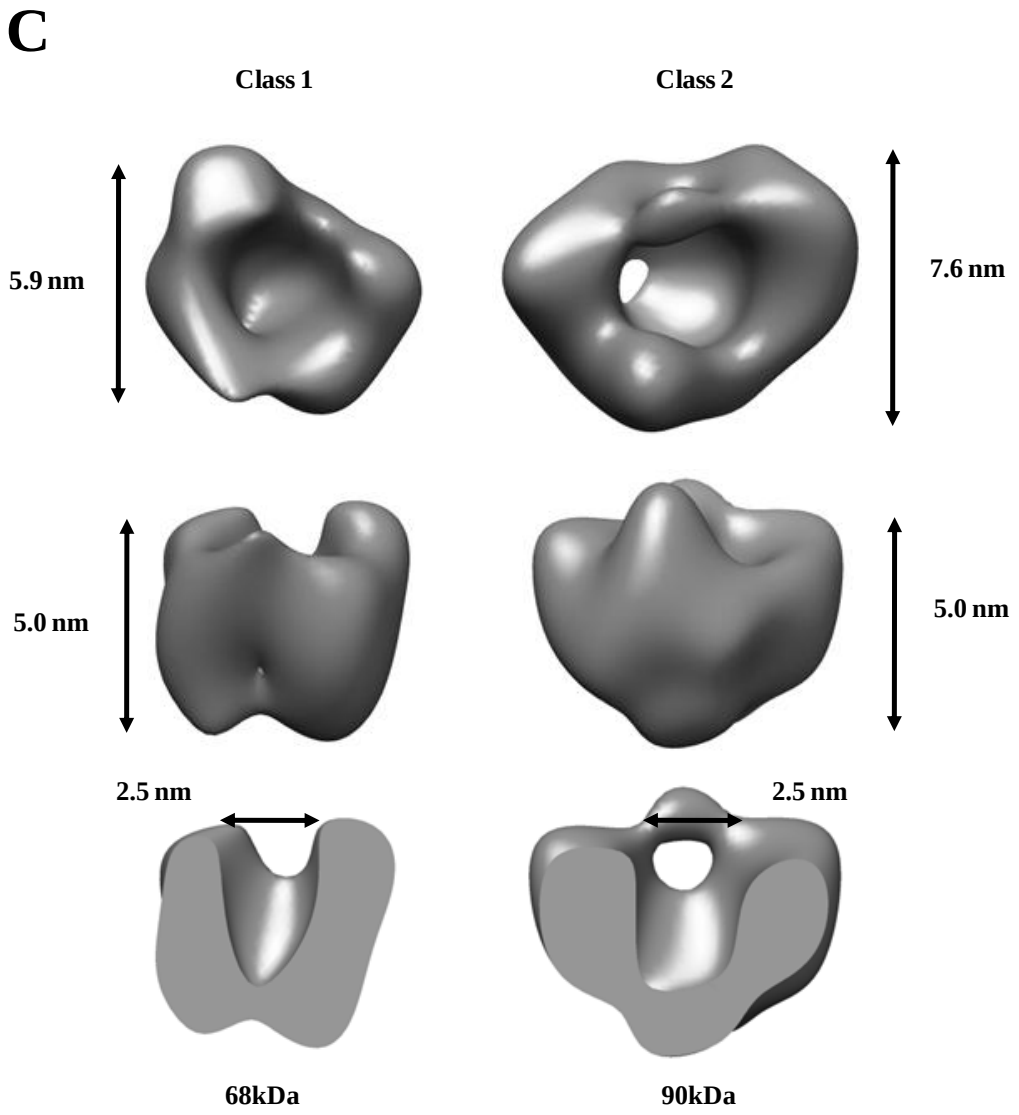
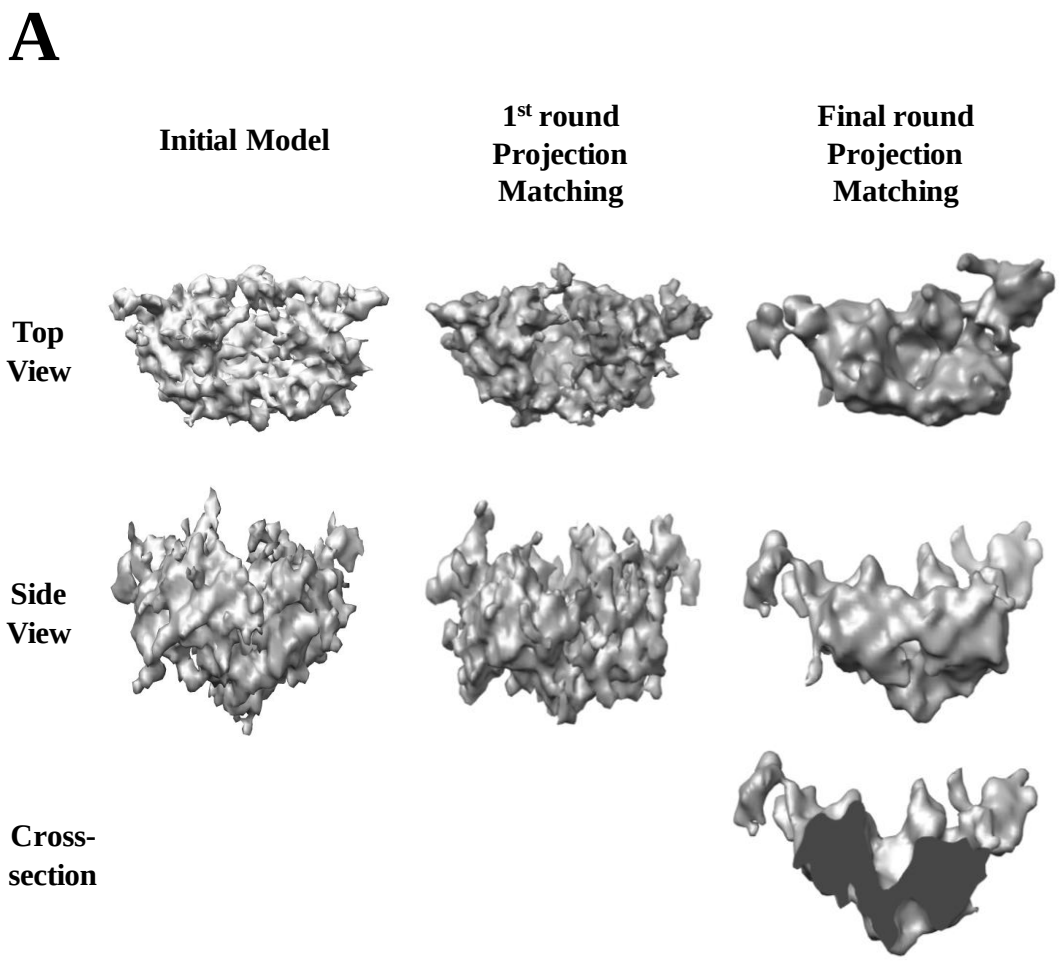



**Figure 4.5.1. Final 3D density maps of TatE-*strep* complex assemblies**
*A+B.* The 2 size classes of single-ringed complexes were refined by projection matching as shown. 3D maps are shown over a number of stages during refinement to show the effects. The presence of a central cavity occluded on one side can be seen in the cross-sectional views. *C.* The final models are shown filtered to 29 Å. and contoured to ~ 5 σ (standard deviations above the mean density). From top to bottom the dimensions shown are: complex diameter, complex height, and 'pore' width. Molecular weights were estimated based on a helical protein packing density of 0.844 Da/Å$^3$. All models were displayed using UCSF Chimera (Pettersen *et al.*, 2004).

### 4.5.2. 3D reconstruction of TatE complex 'Side view'

Finally an attempt was made to produce a relevant 3D reconstruction of the proposed side-view class of TatE complexes. These density maps were generated and refined using the same method used for the top-down reconstruction and the results are shown below in Figure 4.5.2. More particle variance was present in this class compared to the top-down size classes, and therefore the 3D models appear noisier. Throughout the refinement process the presence of a central cavity became clear. A range of low-pass filters were applied to the refined model to accentuate this feature, as shown in Figure 4.5.2.B; multiple cavities, that would link this class to the multi-ringed structures, could not be resolved. These data therefore suggest the rod-shaped structures could represent a flattened population of single ringed structures or a mixed population of single and multi-ringed particles. Although these particles are a structurally heterogeneous minority population they still represent a significant proportion of the TatE particle-set and warrant further investigation in the future.



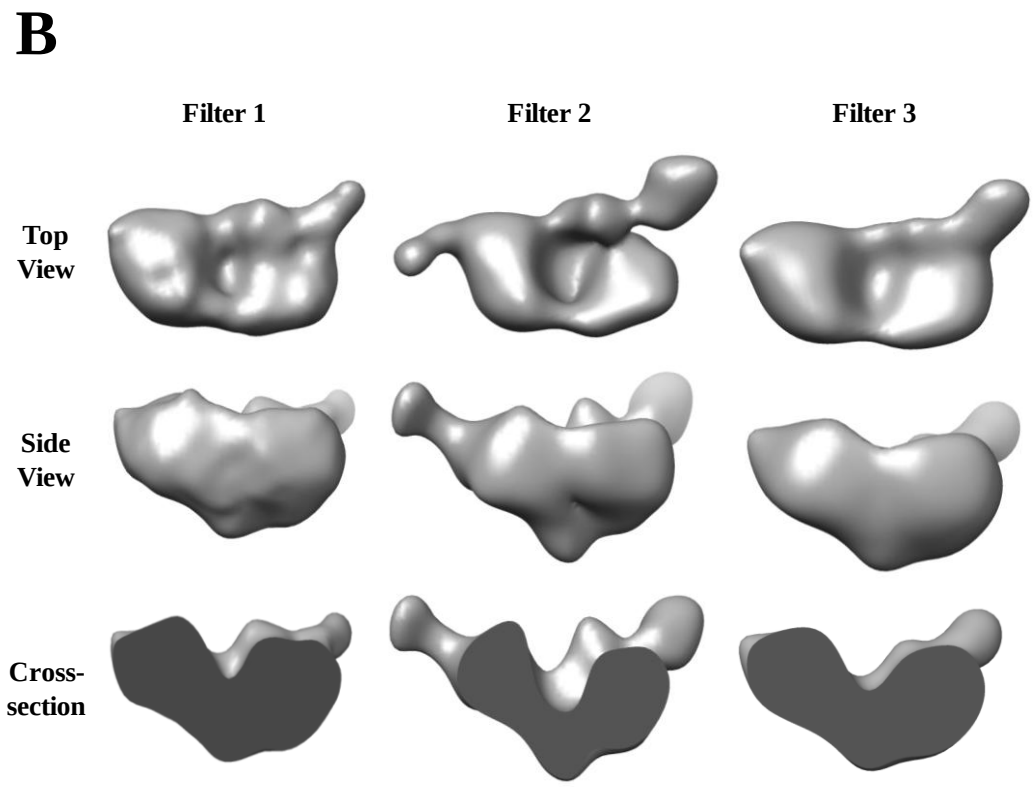



**Figure 4.5.2. 3D reconstructions of TatE complex 'Side' views**
*A.* 3D density maps were generated by projection matching refinement of the side view class of TatE complexes. *B.* Final density maps were generated for the side class by applying a range of low-pass Butterworth filters. In each case a central cleft can be seen that is occluded on one side.

## 4.6. Discussion

Although the role of the Tat(A)BC complex in substrate binding has been demonstrated multiple times using a variety of experimental methods, including cross-linking studies (Alami *et al.*, 2003) and direct EM imaging (Tarry *et al.*, 2009), the function of the TatA complex remains inferred only. The most widely accepted theory suggests that TatA forms a translocation channel made mostly, if not entirely, from TatA subunits. This theory is based heavily on the findings of a previous EM study into *E. coli* TatA complexes where ring-shaped particles of varying size were observed (Gohlke *et al.*, 2005). The data presented here, and published previously (Baglieri *et al.*, 2012), appear inconsistent with this 'size-fitting pore' model.

It has been demonstrated previously that TatE can complement a Δ*tatA* mutant (Sargent *et al.*, 1999) and more recently that it can translocate several Tat substrates including the large TorA protein (90 kDa) (Baglieri *et al.*, 2012). These results indicate that TatE is capable of substituting for *E. coli* TatA. In the previous study of TatA the largest sized class, consisting of 86 particles, measured 13.5 nm across and had a potential channel diameter of 6.5-7.0 nm (Gohlke *et al.*, 2005). However, unlike TatA, TatE presents only 2 major size classes measuring less than 8 nm in diameter, smaller than the smallest sizes TatA complexes. No evidence of particles with a pore 7 nm wide, and therefore potentially capable of accommodating TorA, were seen. As such the small size of these TatE complexes suggests that in this state they would not be capable of forming a pore large enough to translocate folded Tat substrates.

The complexes isolated here by single-particle analysis fit well with the size estimates given by BN-PAGE (see Figure 4.2.2) and it is important to note that larger complexes could not be observed in the whole membranes either, as seen by biochemical analysis, thereby ruling out the possibility of degradation during the purification (Baglieri *et al.*, 2012). Purification in other detergents including digitonin and $C_{12}E_9$ was also not seen to alter the gel filtration elution profile of TatE complexes (see Chapter 8).

The interesting multi-ringed structures suggest some kind of modular assembly intermediate of a larger complex may be present within the TatE sample. Similar structures were observed and discussed previously for TatAd (see chapter 3) and even larger formations were observed for TatAyCy (see chapter 5). These small, modular complexes lend themselves to the idea of a flexible protein conducting channel (PCC), a theory that is currently favourable within the Sec field (Frauenfeld *et al.*, 2011; Haider *et al.*, 2006; Tian and Andricioaei, 2006). Experimenting with the purification conditions to attempt to increase the number of multi-ringed complexes would be an interesting avenue for future research.

A TatBC complex with multiple substrates bound to its periphery has been previously observed by EM (Tarry *et al.*, 2009). A central cavity was shown in this complex; however, it was deemed too small to accommodate a Tat substrate and presented no opening to the particle surface. Further analysis of TatA purified in digitonin presented in this study did not reveal the large central pore shown previously (Gohlke *et al.*, 2005). It is important to note that partial projections of complexes can be produced by variable staining across the grid and this can alter the resolution of such internal cavities. Also, the 3D models of all these Tat complexes have been produced using a random conical tilt reconstruction method. This method results in a cone of missing structural information along the z-axis, which can obscure finer detail within the 3D models. The preferred orientation of these particles on the grid can also obscure whether these cavities are open or closed to the surrounding environment. To circumvent these technical issues related to the imaging method future work may focus on abolishing the preferred orientation. It may be possible to achieve this by experimenting with glow-discharge, staining and gridding procedures, or by using cryo-EM approaches.

Importantly, until a Tat complex can be isolated with clear evidence of a substrate stalled within a central pore or cavity it cannot be safely concluded that these ring-shaped particles produce channels.

# *Chapter 5*

# *Structural investigation of TatAyCy complexes*

## 5.1. Introduction

The TatAyCy pathway is the second minimal (AC-type) Tat system found in *Bacillus subtilis*. The *tatAy* and *tatCy* genes are co-expressed in a single operon and the system is known to translocate a single substrate: YwbN, a heme-containing DyP-type peroxidase (Jongbloed *et al.*, 2004). It is not required for translocation of the TatAdCd substrate PhoD (Jongbloed *et al.*, 2000) and as such represents a separate translocase with differing substrate specificity. The pathway has been termed the 'house-keeping' Tat system of *B. subtilis* as it is constitutively expressed. The A-type component of the *B. subtilis* TatAdCd pathway was been discussed previously in chapter 3. The components of this pathway are only expressed under phosphate limiting conditions.

TatAyCy has been previously shown to form membrane localised complexes estimated to be ~ 200 kDa in size, by gel filtration (Barnett *et al.*, 2009). The disputed cytosolic fraction of TatAy was shown to form very large complexes of ~ 5 MDa that were assumed to be aggregates. In the same study similarities in the complex organisation and biochemical behaviour of both TatAdCd and TatAyCy were noted, both forming relatively small and homogenous complexes leading to the proposal that such features are conserved across Gram-positive AC-type systems.

These findings provide a stark contrast to the highly variable size range of *E. coli* TatA (~100 kDa to over 500 kDa) (Oates *et al.*, 2005) and the proposed pore fitting translocation model (Gohlke *et al.*, 2005). It was also noted that the TatAyCy system, unlike TatAdCd, does not present the ability to translocate as broad a range of substrates (Barnett *et al.*, 2009). This suggests that if any structural differences between Gram-negative and Gram-positive Tat systems are visible they may well be more pronounced in TatAyCy due to its functional rigidity.

Of the currently identified bacterial Tat systems the TatAyCy pathway has received the least attention overall, especially in terms of structural analysis. The TatAdCd complex has been estimated to be 230 kDa as judged by Blue-Native PAGE (Barnett *et al.*, 2008), however attempts with this technique on TatAyCy have so far been unsuccessful, leaving gel filtration estimates as the only current measure of complex size.

The functional importance of a range of conserved residues in the *B. subtilis* TatAy subunit have been investigated in a recent site-specific mutagenesis study (van der Ploeg *et al.*, 2011). The residues mutated in this investigation were selected based on an initial mutagenesis screen in TatAd, which identified a range of N-terminal residues important for the bifunctional properties of the protein (Barnett *et al.*, 2011).

From these experiments on TatAy a substitution of Pro-2, in the extracytoplasmic N-terminal region, with Ala (P2A mutation) was found to produce a stable, but functionally inactive complex, when expressed in either *B. subtilis* or *E. coli* alongside wild-type TatCy (van der Ploeg *et al.*, 2011).

In light of these observations, and aided by the relative homogeneity of complex size shown by the previous biochemical work, a structural investigation was performed to try and visualise the purified TatAyCy complex by single particle EM.

The data presented here show that the formation of small, ringed complexes that vary in size is a property universal to the major bacterial Tat pathways when the proteins are solubilised in detergent.

Indications of higher level modular complex formation are also shown. Additionally, dramatic evidence is presented that demonstrates a major role for the N-terminus of the TatAy in complex formation via a structural investigation of the P2A mutant TatAyCy complex.

All TatAyCy samples were expressed in *B. subtilis* and solubilised membranes were provided by Carmine Monteferrante from the University of Groningen, The Netherlands.

## 5.2. Purification of TatAyCy-*strep* complexes in DDM

### 5.2.1. Affinity purification of TatAyCy-*strep*

A concentrated sample of DDM solubilised membranes (3 ml) was isolated from an 800 ml culture of *B. subtilis* over-expressing Tat AyCy-strep. The sample was added to 4 ml of pre-equilibrated Streptactin™ affinity slurry and left to rotate overnight at 4°C to allow for thorough binding to the entirety of the column. After leaving the column to settle the sample flow-through was eluted before washing with 3 x 8 ml of equilibration buffer to remove non-specifically bound protein. Tightly bound protein was eluted in 12 x 1ml fractions. All fractions, including the initial membranes and column flow-through, were analysed by SDS-PAGE followed by visualisation using silver-stain and immunoblots to the TatCy *strep*-tag II and TatAy protein, as shown in Figure 5.2.1.

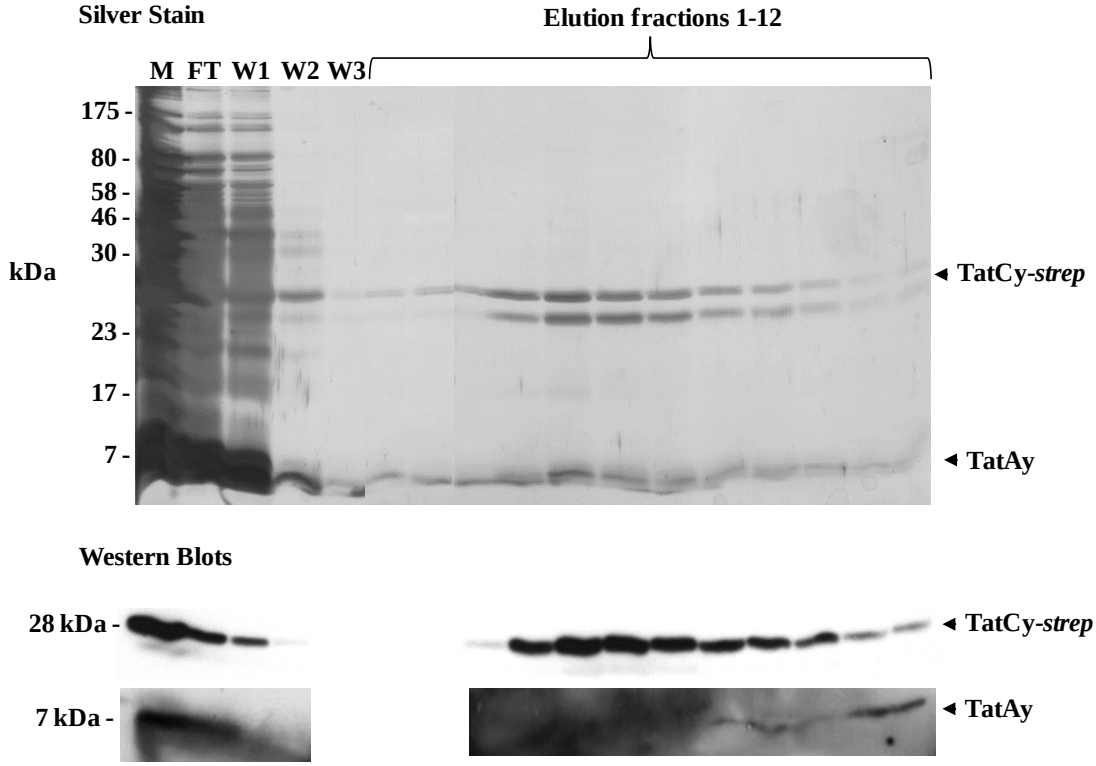



**Figure 5.2.1. Affinity chromatography of TatAyCy-*strep***
Isolated *B. subtilis* membranes were applied to a Streptactin™ column as described above. The whole membranes (M), column flow-through (FT), wash fractions (W1-3) and elution fractions (elution 1-12) were analysed by silver-stain (top panel) and immunoblotted with antibodies against the *StrepII*-tag on TatCy and against TatAy. Positions of TatAy and TatCy are indicated on the right of the figure and molecular weight markers are shown on the left.

As can be seen, some TatCy is lost in the initial flow-through, and presumably has a reduced affinity for the Streptactin™, however the majority is tightly bound to the column and elutes across fractions 3-12. Some TatAy is also lost during the initial flow-through, but a significant amount appears to co-elute with TatCy up to elution fraction 12. The silver-stain shows the elution fractions to be very pure with the exception of a faster migrating band that mirrors TatCy. This lower band most likely represents a proteolytic cleavage product of TatCy where the C-terminal *strep*-tag II has been removed. As such this band is not evident on the anti-strep western blot. These data suggest that although a proportion of the Tat protein is lost using this rotation-incubation method, the final eluted sample appears to be very pure.

### 5.2.2. Gel filtration chromatography of TatAyCy-*strep*

All Tat containing elution fractions (3-12) were pooled and concentrated ~ 10x using a 30 kDa MWCO Vivaspin™ concentrator, to give a final concentrate of ~ 1 ml. A 240 µl sample of the concentrate was loaded onto a Superdex-200 gel filtration column and the elution fractions across the resolving range of the column were again analysed by SDS-PAGE. As shown in Figure 5.2.2.A. TatCy primarily elutes across fractions 18-26 with the peak of TatCy elution corresponding to ~ 166 kDa using standard soluble molecular weight marker proteins. Interestingly, TatAy is shown to elute tightly across fractions 23-27, peaking at ~ 100 kDa. Densitometry of the corresponding Western blots (Figure 5.2.2.B.) gives a clearer picture of this elution profile in which there is a lack of correlation between the elution peak of TatCy and TatAy. As this TatAy was bound tightly to the affinity column and co-eluted with TatCy it can be assumed that a TatAyCy complex is present, however this disparity in size elution profiles between the two proteins suggests two potential scenarios. Either a dynamic range of complexes may be present with differing ratios of TatAy to TatCy, or the TatAyCy complex is unstable with some TatAy detaching. It is important to note that TatCy is present in all TatAy containing fractions, as such all fractions can be said to contain potential TatAyCy complexes which vary in size.

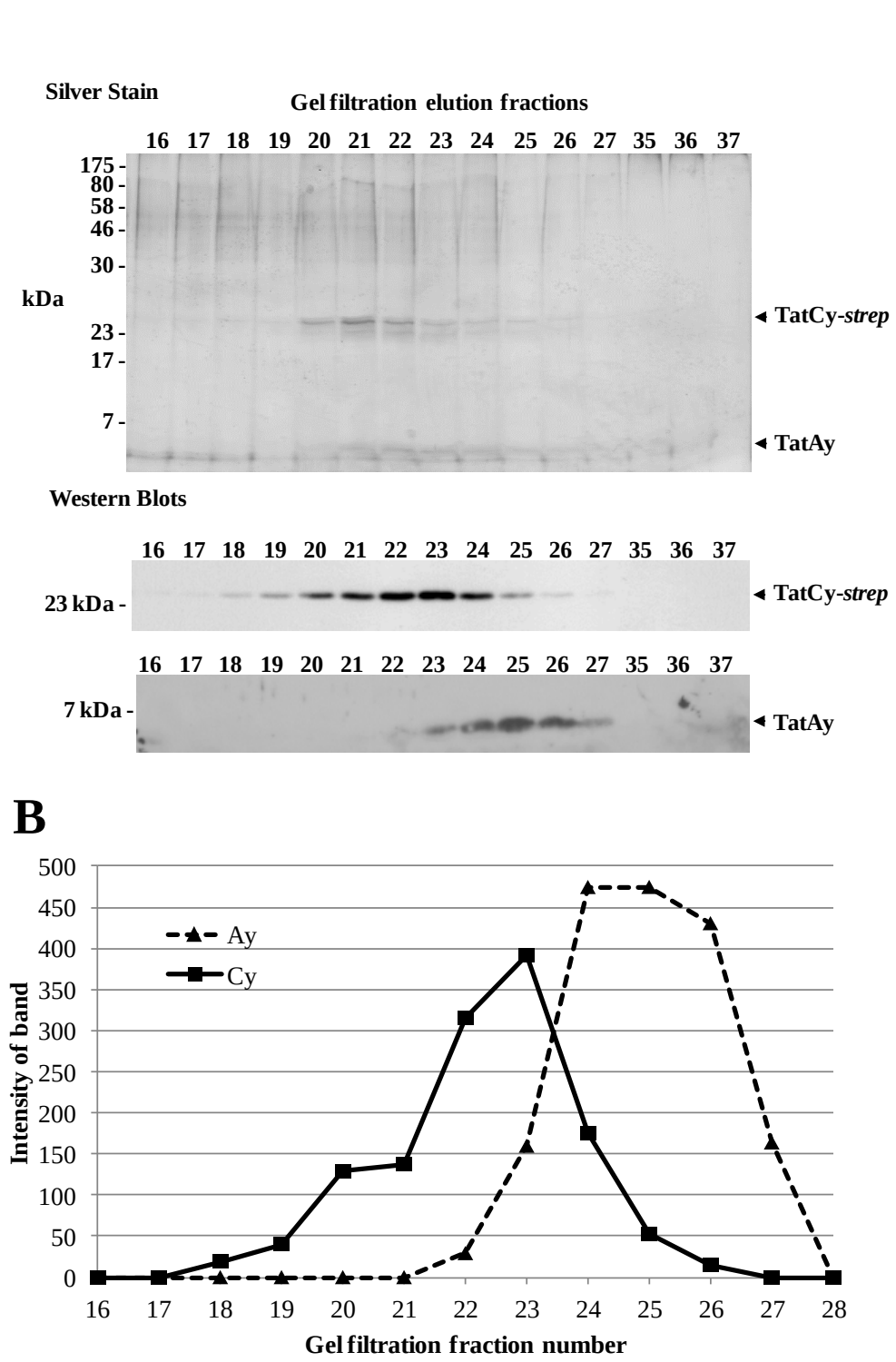



**Figure 5.2.2. Gel filtration and densitometry of TatAyCy-*strep***
*A.* 240 µl of TatAyCy-*strep* concentrate was applied to an equilibrated Superdex™ 200 gel filtration column as described above. Elution fractions corresponding to the separation range of the column (16-27) and covering the bulk flow (35-37) were analysed by silver-stain gel and immunoblotted with antibodies against the *StrepII*-tag on TatCy and against TatAy. Positions of TatAy and TatCy are indicated on the right of the figure and molecular weight markers are shown on the left. *B.* Densitometric analysis of the corresponding Western-blots shown in *A.*

## 5.3. Electron microscopy of TatAyCy-*strep*

### 5.3.1. TatAyCy forms small, ringed complexes

Samples were taken from gel filtration fractions covering the peak of Tat protein elution (22-26). Initial negative stain grids were prepared using undiluted aliquots to avoid the potential introduction of artefacts during dilution. The standard grid preparation method was used. As shown in Figure 5.3.1 the majority of particles appear small and round with a central pool of stain, resembling those seen in both the TatAd and TatE samples. Particles range in size from ~ 6 - 9.5 nm with the majority being ~ 8 nm in diameter.

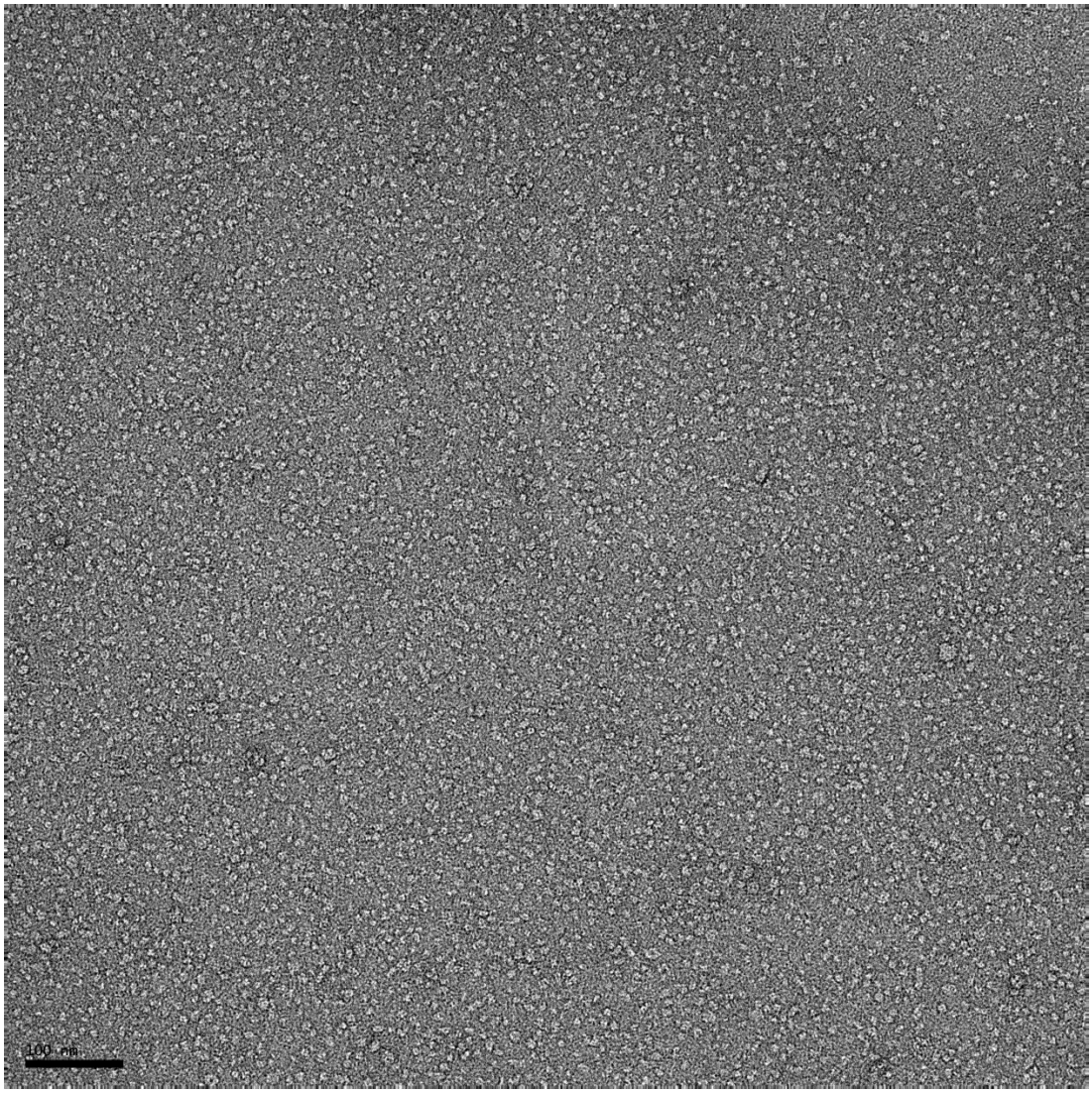



**Figure 5.3.1. Micrograph of TatAyCy-*Strep* undiluted gel filtration fraction 23**
The micrograph was taken at ~57,000x magnification under ~1.5 µm defocus. The grid was stained with 2% uranyl acetate. TatAyCy forms a series of ringed particles varying in size from ~ 6.0 - 9.5 nm in diameter.

To improve particle separation the sample was diluted 1:5 using buffer + 0.02% DDM as for the TatAd sample (see chapter 4). As shown in Figure 5.3.2 the particle density is dramatically reduced, but the morphology of the particles is unaffected. Evidence of double-ringed structures was also observed; see Figure 5.3.3, despite the dilution, similar to those shown previously for the TatE sample (see chapter 5). This indicates that TatAyCy may be capable of modular multimerisation.

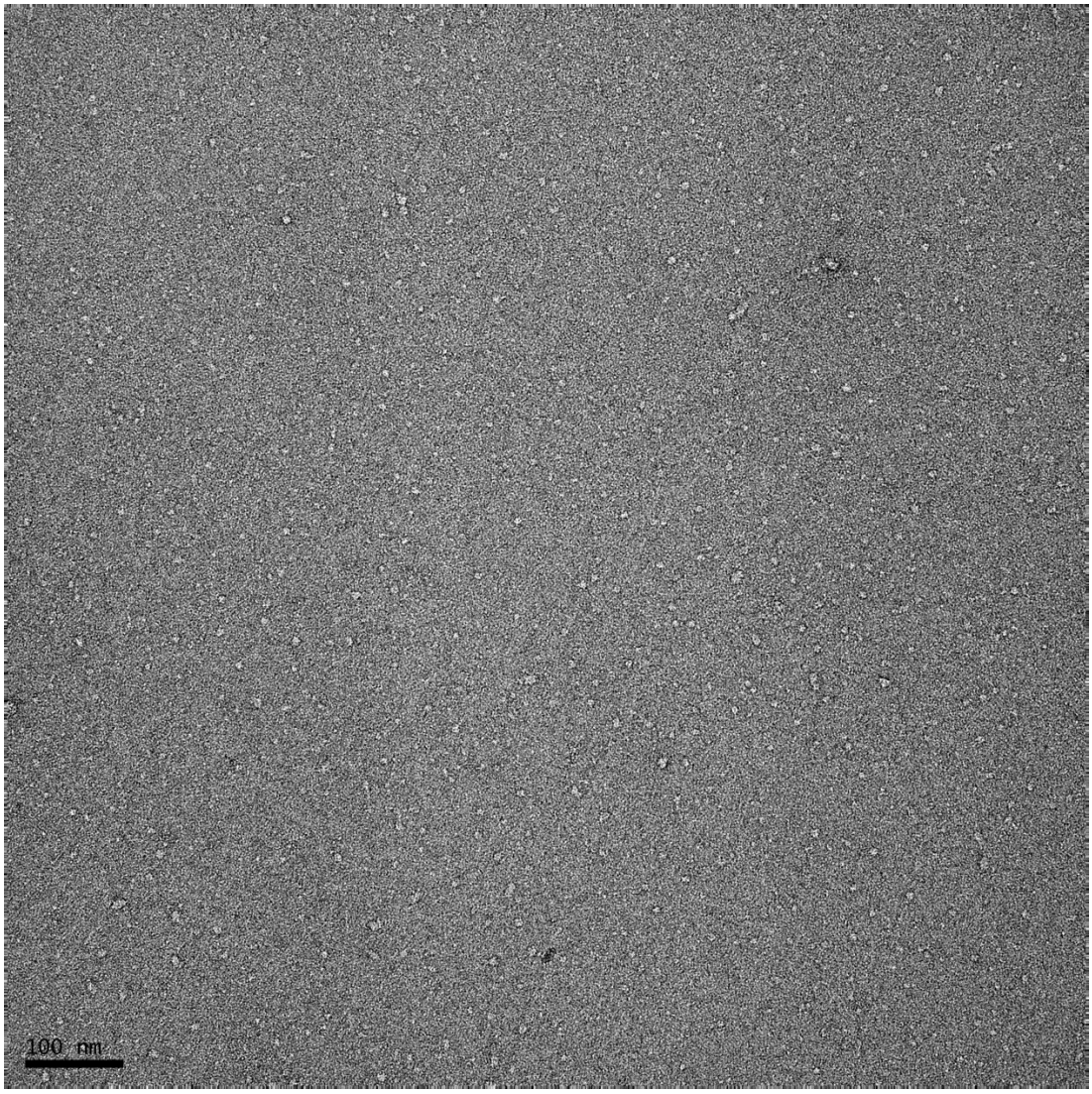



**Figure 5.3.2. TatAyCy-*Strep* GF 23 1:5 dilution in GF buffer + 0.02% DDM**
The micrograph was taken at ~57,000x magnification under ~1.5 µm defocus. The grid was stained with 2% uranyl acetate.

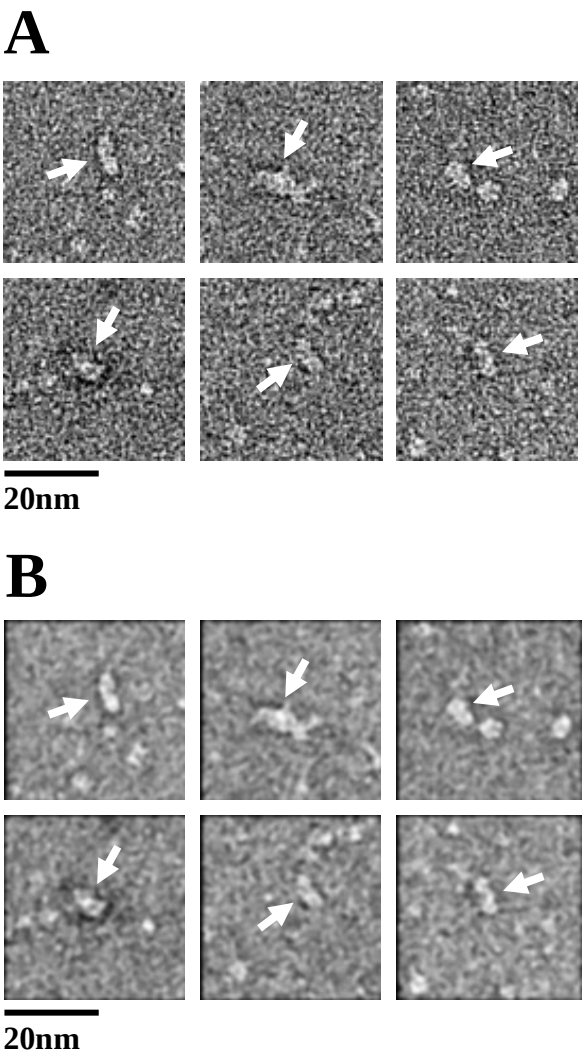



**Figure 5.3.3. Evidence of multi-ringed structures within the TatAyCy sample**
*A.* Examples of double-ringed particles seen within the TatAyCy-*strep* sample are indicated by white arrows. A central pool of stain can be seen within each ring. These particles measure ~ 10 nm long and ~ 5 nm wide. *B.* Median filter applied to the same images as shown in *A.*

EM evaluation of gel filtration fractions 22 and 26 also displayed small ring shaped particles and did not reveal any discernible correlation between complex size and relative abundance of Tat subunits. Overall these data indicate that the Tat proteins are able to form stable, low order structures but the formation of TatAyCy complexes appears to be unstable under these conditions. The interaction between TatAy and TatCy is clearly evident from the biochemical data, but this interaction appears to be transient or unstable in nature.

### 5.3.2. Higher order tetrameric assemblies of TatAyCy

When analysing gel filtration fraction 24 an interesting subset of particles were identified. As shown in Figure 5.3.4 small, ringed particles were again observed with no clear change in morphology from those of fraction 23. However, amongst these particles larger ringed structures with a striking tetrameric arrangement were also seen.

These particles were found infrequently (1 or 2 per micrograph at best) along with other large, but less defined structures, of a similar size; these potentially represent broken and flatten copies of the same particles. Examples of these larger structures are shown clearly in Figure 5.3.5. These well defined complexes consist of 4 copies of small ringed particles each measuring ~ 6 - 7 nm in diameter, the full tetramer is ~ 17 nm wide with a central stain pool of ~ 5.5 nm wide. This central 'pore' is larger than those seen for any of the TatAd or TatE particles.

Based on the purity of this sample (see Figure 5.2.1) and the similarity in size and shape of the component densities to the individual ringed particles, these tetramers appear to represent an unstable but highly ordered complex of Tat proteins. The relative contribution of subunits within the structure is unclear but they presumably consist of both TatAy and TatCy subunits.

These tetramers were not identified in any Tat samples where only the A (or A like) subunit is present and as such these data fit well with the observation that higher order oligomerisation of TatA normally requires the presence of the TatBC subunits (Leake *et al.*, 2008). These particles are larger than other identified Tat complexes, are only found in the presence of all subunits required for translocation and are found in low abundance, which is indicative of an inherently unstable structure or a transient coalescence of subunits. Taken together these finding indicate that these tetramers may be representing a higher order assembly of the TatAyCy complex that could be of interest for future studies. However, until they can be isolated in greater numbers the biological significance of these assemblies cannot be verified.

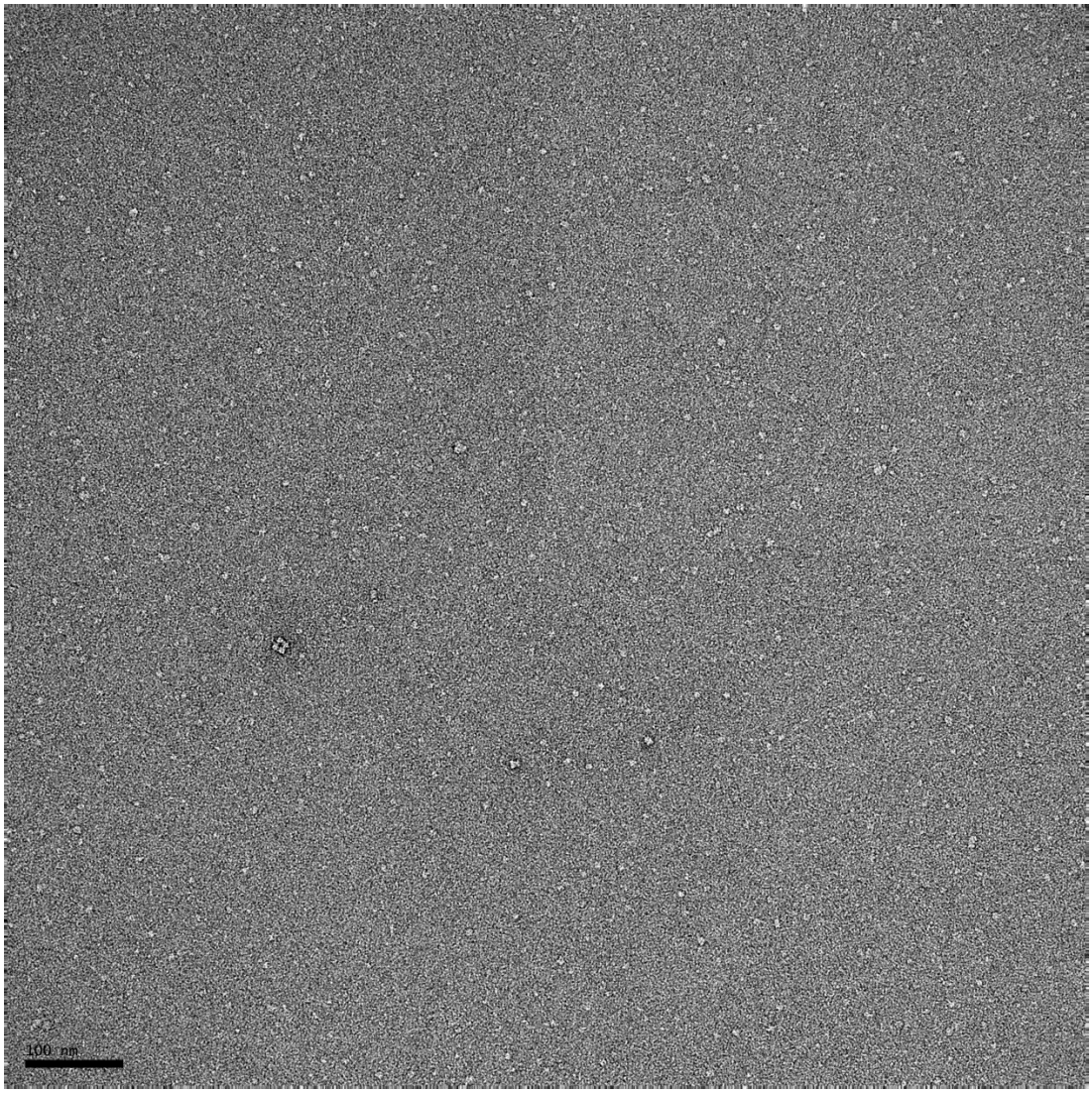



**Figure 5.3.4. TatAyCy-*Strep* GF 24 1:5 dilution in GF buffer + 0.02% DDM**
The micrograph was taken at ~57,000x magnification under ~1.5 µm defocus. The grid was stained with 2% uranyl acetate.

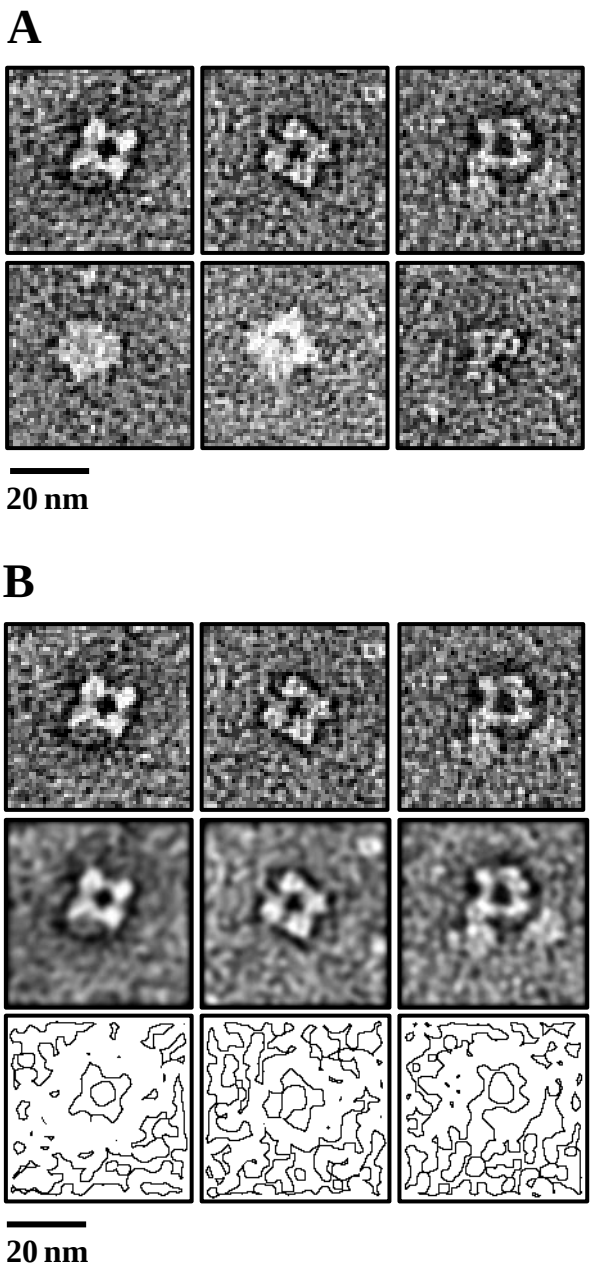



**Figure 5.3.5. Potential tetrameric complexes observed for TatAyCy-*strep***
*A.* Montage of large multimeric complexes of Tat protein. The top panels show clear tetramers of small ringed units, each ~ 5.5-7 nm wide, joined together by short extensions. A pool of stain is clearly visible within most of these small units. The central stain pool of the tetramer is ~ 5.5 nm wide. The lower panels show examples of distorted or broken complexes. *B.* The same particles shown in the top panel of *A.* are displayed with a median filter (middle panel) and as a binary outline image (lower panel) to clearly identify the modular structure.

## 5.4. Purification of TatAyCy-*his* wild-type vs. P2A mutant

The results for the complex formation of the strep-tagged sample were interesting but unexpected. In order to further investigate these complexes by EM and to conduct Nanogold binding experiments a construct with a C-terminal Hexahistidine tag was used.

It has previously been shown that a range of single amino acid mutations in TatAy can abolish translocation of the substrate YwbN by TatAyCy in *B. subtilis*. (van der Ploeg *et al.*, 2011). These mutations were shown to have variable effects on complex stability as seen by gel filtration but no further structural investigation was undertaken. Here the effects of substituting Pro-2 in the extracytoplasmic N-terminal region with Ala (P2A mutation) on complex formation and structure are shown. For these purifications larger samples of solubilised membranes (~ 11 ml) were provided, and a larger affinity column was used (10 ml Talon ™). All other conditions and reagents used were the same as used for the *strep*-tagged sample.

### 5.4.1. Affinity purification of wild-type TatAyCy-*his*

SDS-PAGE gel analysis was performed as before with visualisation using Instant Blue™ stain and immunoblots against the His-tag of TatCy-*his* and TatAy. Figure 5.4.1. shows the results of the affinity column purification of TatAyCy-*his* wild-type. These results are comparable to those of the *Strep*-tagged sample, with clear bands present for TatCy and TatAy with little contaminating proteins. Once again a lower migrating band of TatCy can be seen that most likely corresponds to cleavage of the C-terminal tag.

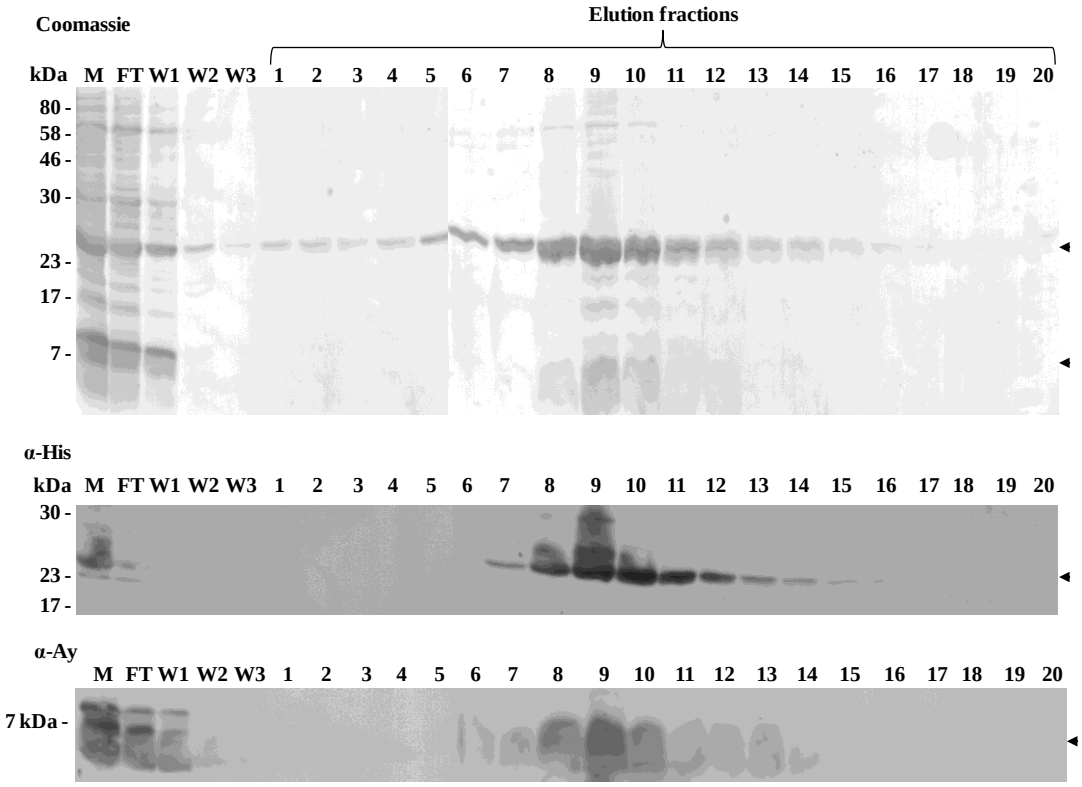



**Figure 5.4.1. Affinity chromatography of wild-type TatAyCy-*his***
Isolated *B. subtilis* membranes were applied to a Talon™ column as described above. The whole membranes (M), column flow-through (FT), wash fractions (W1-3) and elution fractions (elution 1-20) were analysed by Coomassie stain (top panel) and immunoblotted with antibodies against the *his*-tag on TatCy and against TatAy. Positions of TatAy and TatCy are indicated on the right of the figure and molecular weight markers are shown on the left. Compare with Figure 5.4.2. below.

### 5.4.2. Affinity purification of P2A mutant TatAyCy-*his*

The results of the P2A mutant purification are shown in Figure 5.4.2. These appear very similar to the wild-type in terms of purity and clipping of the TatCy band is again evident, however much higher amounts of TatAy are detected on the stained gel and Western blot.

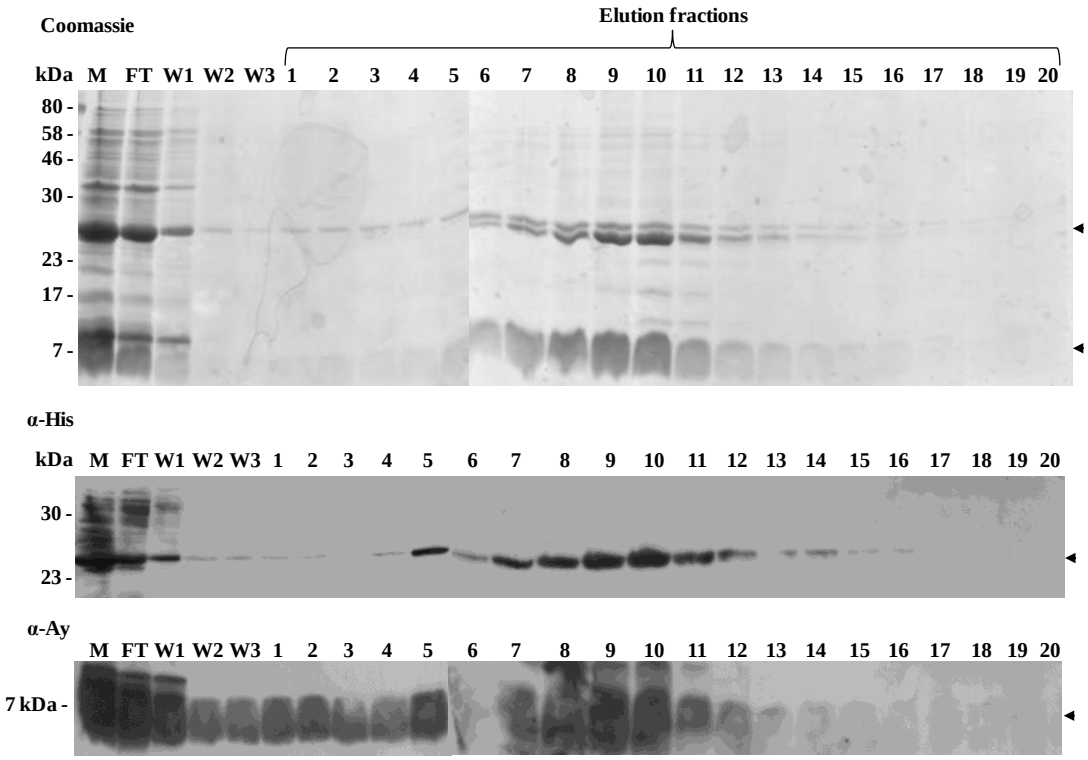



**Figure 5.4.2. Affinity chromatography of mutant TatAyCy-*his***
Isolated *B. subtilis* membranes were applied to a Talon™ column as described above. The whole membranes (M), column flow-through (FT), wash fractions (W1-3) and elution fractions (elution 1-20) were analysed by Coomassie stain (top panel) and immunoblotted with antibodies against the *his*-tag on TatCy and against TatAy. Positions of TatAy and TatCy are indicated on the right of the figure and molecular weight markers are shown on the left. Compare with Figure 5.4.1. above.

The total protein concentration for both samples across all fractions was determined by Bradford's assay. As shown in Figure 5.4.3. the total protein concentration is very similar for both membrane samples: wild-type ~ 929 µg/ml, Mutant ~ 1020 µg/ml. Taken together these results suggest overall more TatAy is present relative to TatCy in the mutant than in the wild-type. However, more TatAy is lost during the washes

in the mutant sample and is present in the initial elution fractions where little TatCy can be detected. This indicates that this excess TatAy is not all associated with TatCy. Again proteolytic cleavage of the C-terminal tag is evident for TatCy, in all 3 samples both clipped and unclipped bands are seen in the affinity column elution fractions indicating that these TatAyCy complexes contain both forms of TatCy.

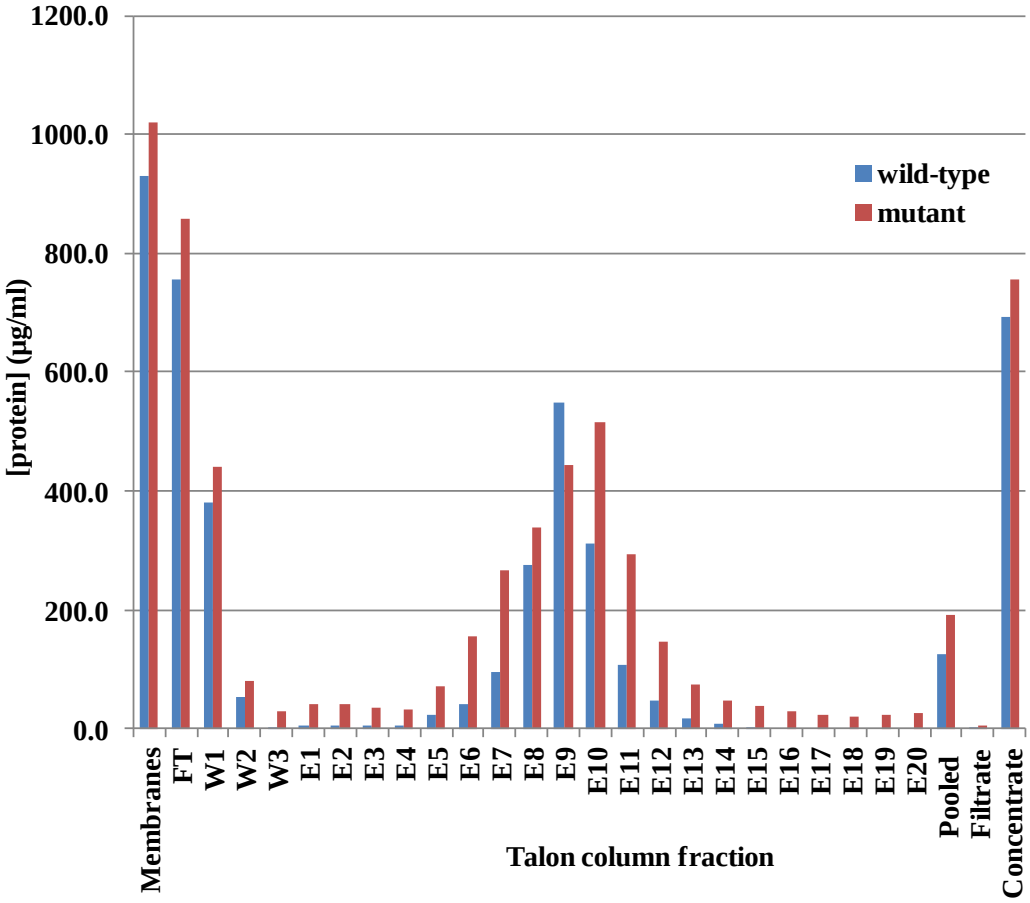



**Figure 5.4.3. Protein concentration across purified TatAyCy-*his* fractions**
Protein concentration was estimated using Bradford's reagent and a BSA standard curve.

### 5.4.3. Concentration of wild-type vs. mutant TatAyCy-*his*

Tat-containing elution fractions were pooled and concentrated as before, and SDS-PAGE gels were performed to highlight the difference between the wild-type and mutant samples (Figure 5.4.4.). These gels clearly show the increased amount of TatAy present in the mutant sample, whereas the TatCy-*his* bands appear comparable. It is also worth noting that no Tat proteins are visible in the filtrate for

either sample. Since both the Tat subunits are below the MWCO of the concentrator (30 kDa) it can be inferred that all Tat protein is present in the form of multimeric complexes.

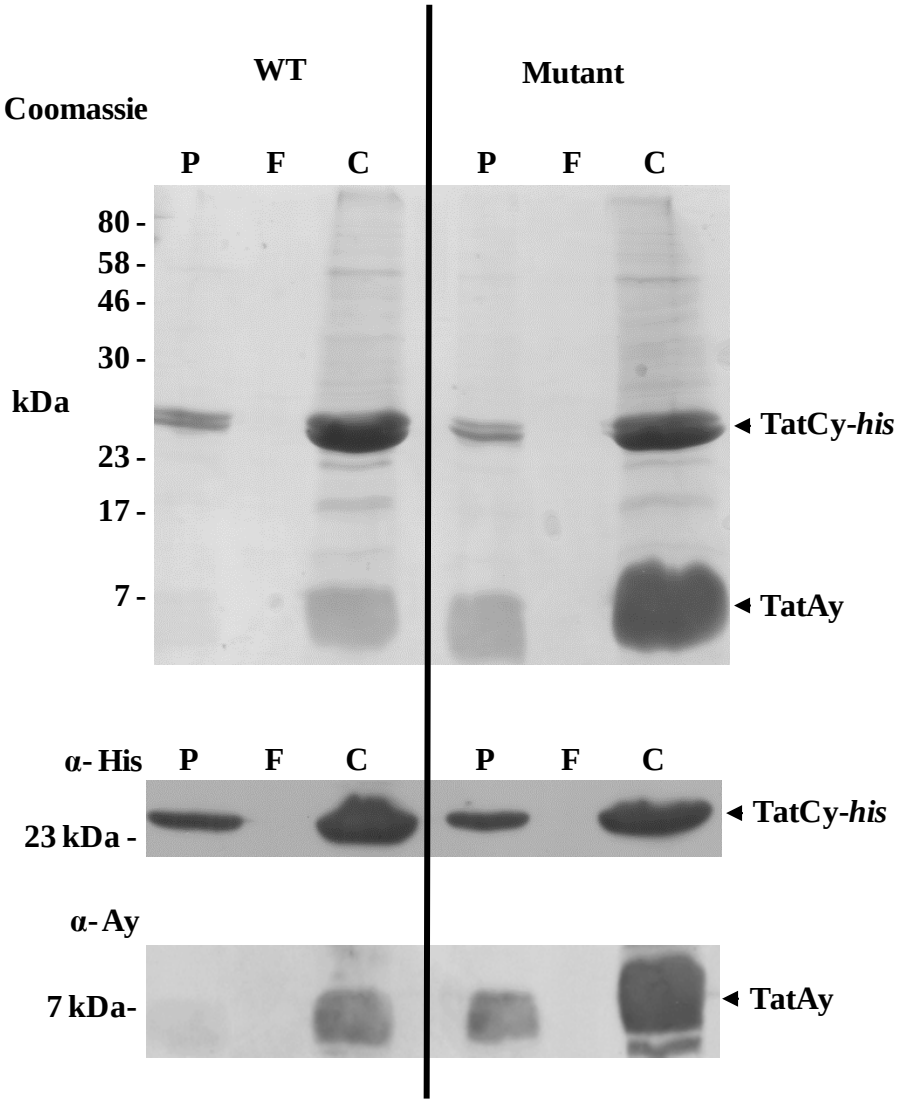



**Figure 5.4.4. Concentration of wild-type and mutant TatAyCy-*his***
Tat containing Talon™ column elution fractions were pooled for the wild-type (fractions 7-16) and mutant (fractions 5-14) before concentrating 10x using Sartorius Stedim Vivaspin 20 concentrators (30 kDa MWCO). The pooled protein (P), filtrate (F) and concentrate (C) were analysed by silver-stain and immunblotted with antibodies against the *his*-tag on TatCy and against TatAy. Positions of TatCy-*his* and TatAy are indicated on the right of the figure and molecular weight markers are shown on the left.

### 5.4.4. Gel filtration chromatography of wild-type vs. mutant TatAyCy-*his*

Analysis of these complexes by gel filtration reveals a striking difference in size distribution of the Tat subunits between the two samples. Figure 5.4.5. shows the UV trace over both gel filtration runs along with the elution volumes of soluble protein standards. As with all detergent solubilised membrane proteins the molecular weights given by such a standard curve are rough estimates only due the effect of the detergent. From this trace a degree of aggregation can be seen for both samples due to the large peak at the void volume. This is followed by a broad peak over the separation range of the column, peaking at ~ 290 kDa for the wild-type and shifting up to ~ 560 kDa for the mutant. In both cases a small peak is seen at ~ 84 kDa.

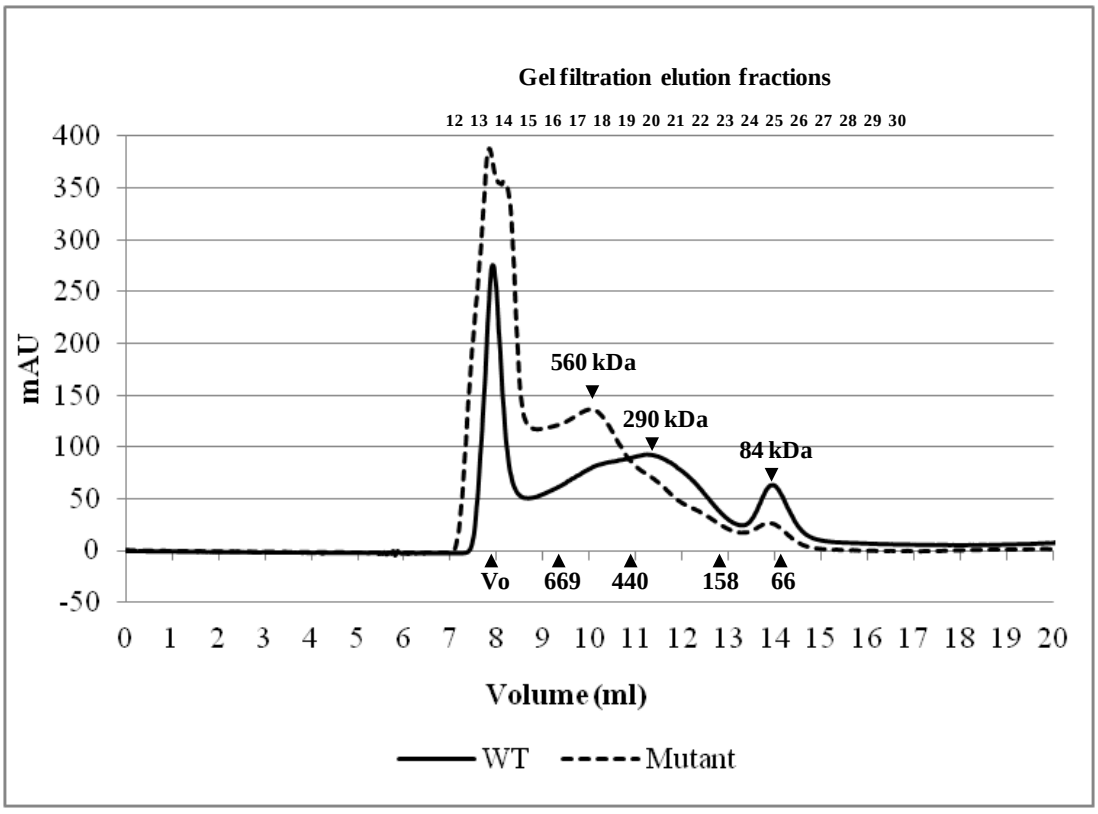



**Figure 5.4.5. Gel filtration profile of wild-type vs. mutant TatAyCy-*his***
The 220 nm UV trace of wild-type and P2A mutant TatAyCy-*his* gel filtration runs. Elution volumes for protein standards are shown along with estimated molecular weights for the sample peaks.

SDS-PAGE analysis of the gel filtration fractions covering this elution profile (7-15 ml) reveal further differences between the wild-type and the P2A mutant. The gels for the wild-type are shown in Figure 5.4.6. along with densitometry of the Western blots (figure 9B). From the Coomassie stain TatCy-*his* can be seen to elute across all fractions with relative peaks in fractions 14, 21 and 26, corresponding to the peaks seen on the UV trace. A small amount of TatAy, barely detectable by Coomassie stain but clearly shown in the Western blot, co-elutes with TatCy-*his* across fractions 21-26. Interestingly the Western blot against the His-tag on TatCy-*his* does not correlate with the band intensity seen by Coomassie stain covering fractions 25-30. This suggests that these later fractions contain almost exclusively the clipped TatCy, the small size difference in which has not been properly resolved by the Coomassie stained gel.

The same analysis of the P2A mutant is shown in Figure 5.4.7. Here, strikingly large amounts of TatAy can be clearly seen in the Coomassie stain as well as the corresponding Western blot. The elution profile of TatCy-*his* appears to be unchanged from the wild-type, but the vast majority of this 'excess' TatAy appears to elute in the void volume (fractions 13+14) trailing into the wild-type elution profile (fractions 21-26). It therefore seems that this single point mutation is causing the formation of excessive TatAy complexes or aggregates (>600 kDa) with little associated TatCy-*his*. However the presence of mutant TatAy across the wild-type elution profile fractions suggests that TatAyCy-*his* complexes are still able to form. The contrast between the elution profiles of wild-type and mutant TatAy is shown clearly in Figure 5.4.8. These data indicate that the formation of these excessive TatAy aggregates may be the cause of the loss of translocation activity.

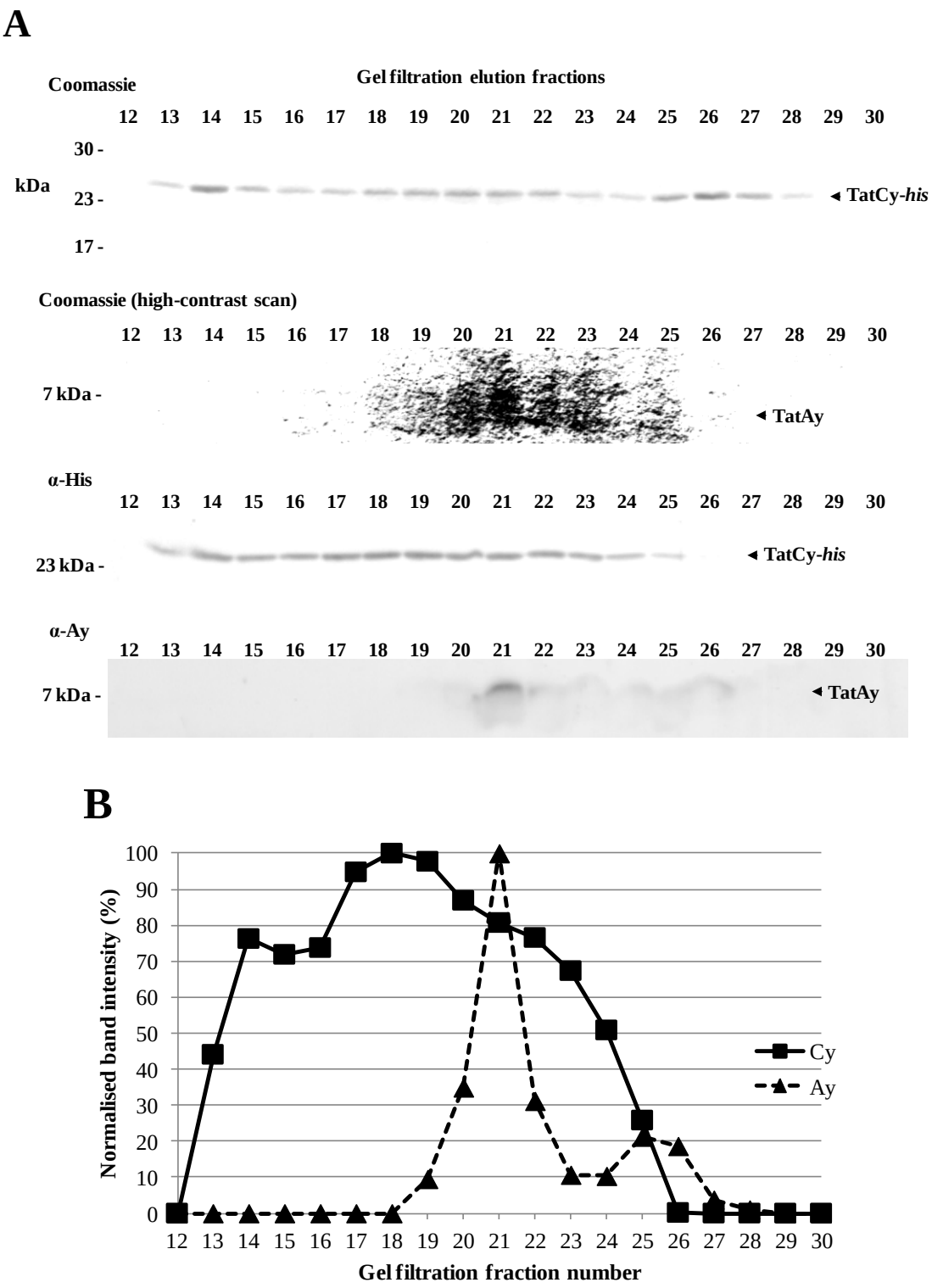



**Figure 5.4.6. Analysis of the wild-type TatAyCy-*his* gel filtration profile**
*A.* SDS-PAGE gel analysis across the gel filtration elution profile is shown. The Coomassie stained gel is split into two sections. The top one shows TatCy-*his* and the lower one is shown at very high contrast due to the low staining of the TatAy present across fractions 19-26. The corresponding Western blots are shown below the Coomassie stains. *B.* Densitometric profile of the Western blot bands shown in *A*, the normalised intensity of each band is plotted. Compare with Figure 5.4.7 below.

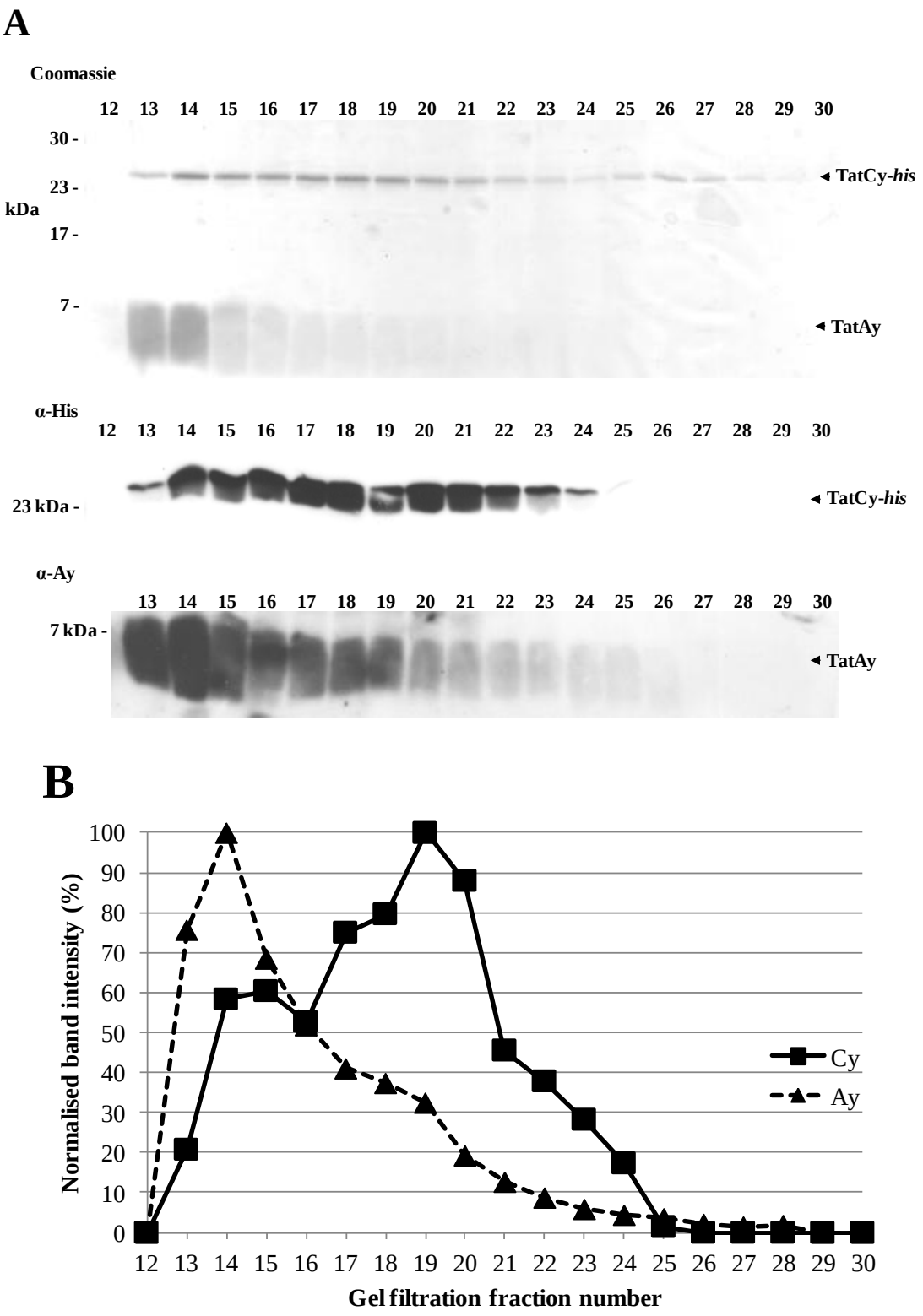



**Figure 5.4.7. Analysis of the mutant TatAyCy-*his* gel filtration profile**
*A.* SDS-PAGE gel analysis across the gel filtration elution profile is shown. The Coomassie stain shows TatCy-*his* (upper band) and TatAy (lower band). The corresponding Western blots are shown below the Coomassie stain. *B.* Densitometric profile of the Western blot bands shown in *A*, the normalised intensity of each band is plotted. Compare with Figure 5.4.6 above.

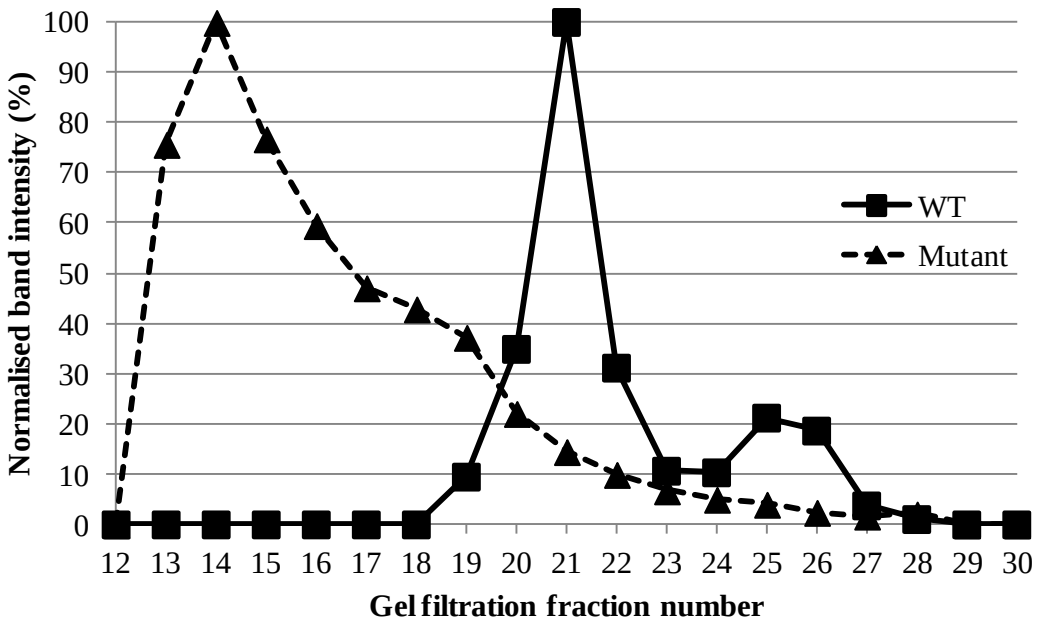



**Figure 5.4.8. Comparison of the wild-type and mutant TatAy elution profiles**
The densitometric profiles of the Western blots for TatAy Wild-type and P2A mutant, as shown in Figures 5.4.7 and 5.4.6, are plotted together for comparison.

## 5.5. Electron microscopy of TatAyCy-*his* wild-type vs. P2A mutant

To investigate the effect of the P2A mutation on TatAyCy complex formation samples were prepared for single particle imaging under the electron microscope. Samples were taken from gel filtration fractions 14 and 21 for the wild-type and mutant, corresponding to the aggregate (void) and wild-type peak fraction respectively. Initial negative stain grids were prepared using undiluted aliquots to avoid the potential introduction of artefacts during dilution. Standard grid preparation method was used.

### 5.5.1. Single-particle EM of wild-type TatAyCy-*his* void fraction

As shown in Figure 5.5.1. the wild-type void fraction consists of a wide variety of irregular amorphous particles that vary greatly in both shape and size; particles ranging from ~8 nm to over 100 nm are evident. The smallest of these particles form a background of round and relatively homogenous structures. As evident from the biochemical data shown earlier (Figure 5.4.6.) this fraction consists of TatCy-*his* with no detectable TatAy or contaminants. Therefore it can be concluded that TatCy-*his* is able coalesce into a wide range of shapes and sizes in the absence of sufficient TatAy and that such complexes are structurally unstable at least under these purification conditions. If such complexes were stable all visible particles present in this gel filtration fraction would be expected to be over the 600 kDa size limit of the Superdex 200 column.

Figure 5.5.2. shows a 1:4 dilution of this fraction, prepared using gel filtration buffer + 0.02% DDM. Although the relative number of particles has been reduced the size distribution and morphology of the particles remains unaffected, showing that the large aggregates are not caused by overcrowding of the grid and that they cannot be disaggregated by simple dilution.

To provide further confirmation as to the constitution of these aggregates a grid was prepared with the addition of 1.8 nm Ni-NTA-Nanogold® (Nanoprobes). A fresh grid was prepared using the same 1:4 dilution of TatAyCy but the 1 minute incubation of the sample on the grid was followed by a 10 minute incubation with a 1:10 dilution of Nanogold (diluted in GF buffer minus detergent), before thoroughly washing and

staining (see methods). From Figure 5.5.3.A multiple intense black dots of 1.8 nm diameter can clearly be seen clustering over the large aggregates indicating that multiple his-tagged proteins are surface accessible. The Nanogold can be seen more clearly in the inset images shown in Figure 5.5.3.B+C. These results provided clear evidence that the aggregates consist of TatCy-*his*.

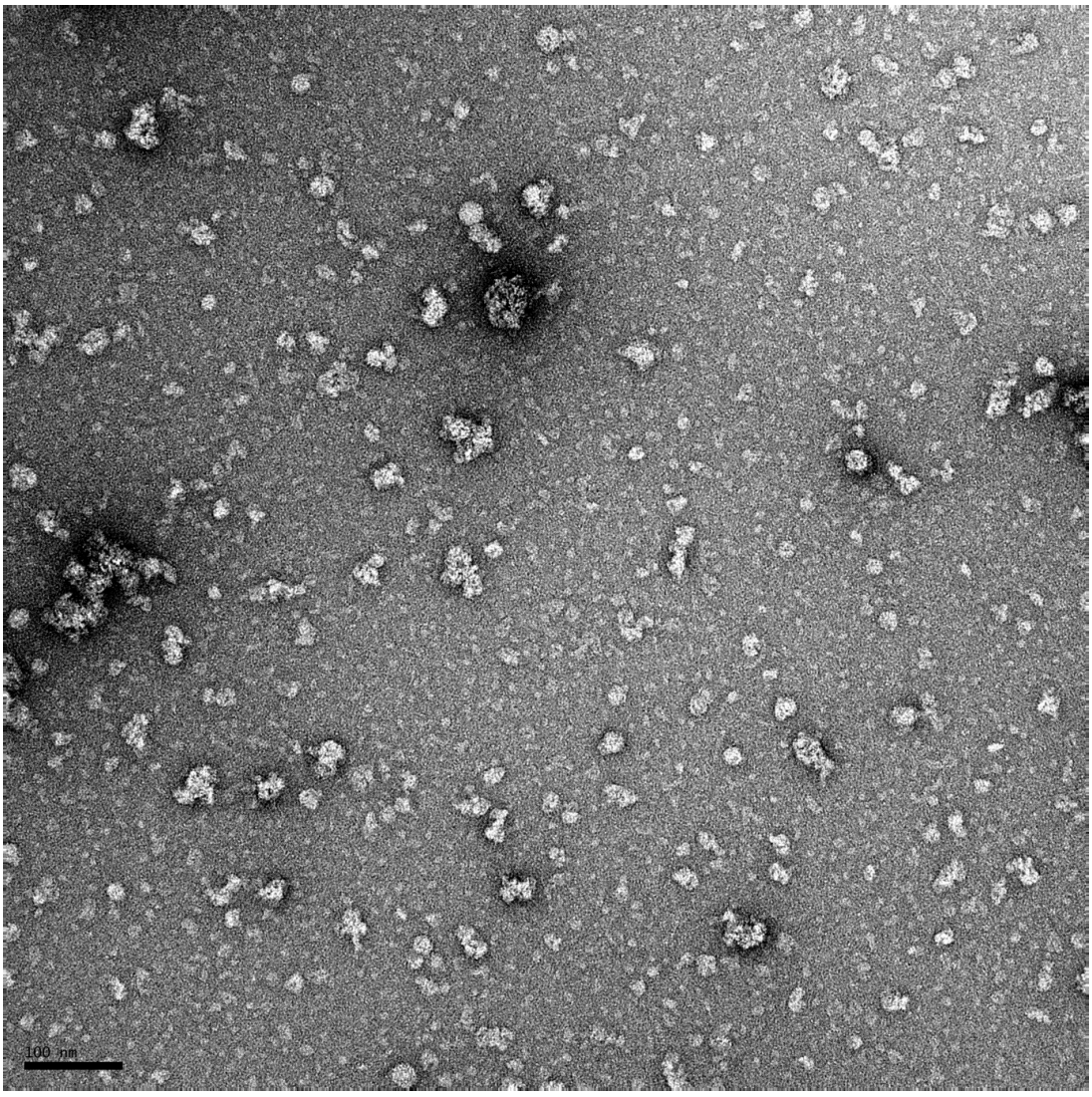



**Figure 5.5.1. Micrograph of wild-type TatAyCy-*his* void fraction undiluted**
The micrograph was taken at ~57,000x magnification under ~1.5 µm defocus. The grid was stained with 2% uranyl acetate. TatAyCy forms a wide variety of irregular amorphous particles that vary greatly in both size and shape (~8 nm to over 100 nm).

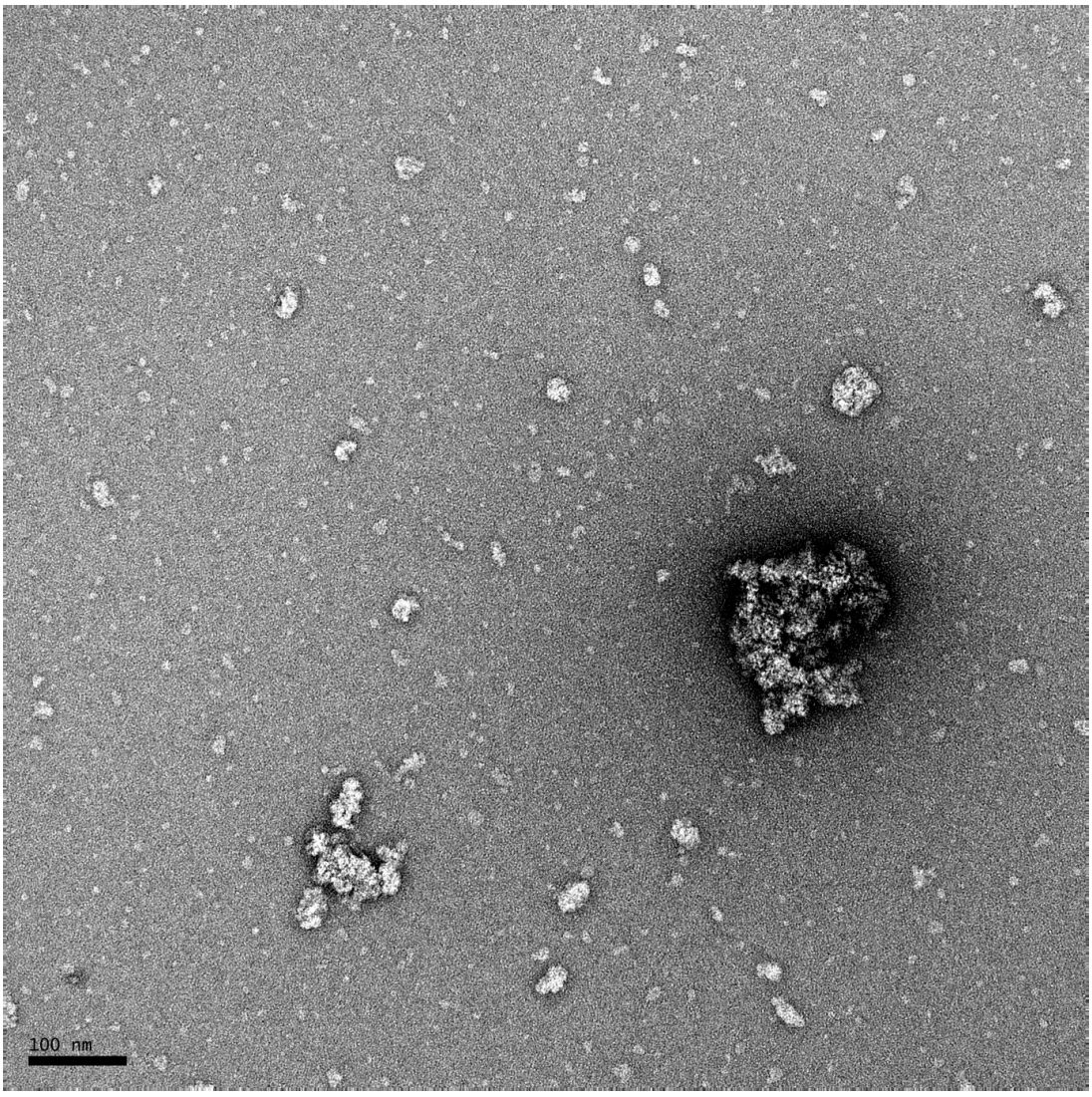



**Figure 5.5.2. Micrograph of wild-type TatAyCy-*his* void fraction 1:4 dilution**
The micrograph was taken at ~57,000x magnification under ~1.5 µm defocus. The grid was stained with 2% uranyl acetate. TatAyCy forms a wide variety of irregular amorphous particles that vary greatly in both size and shape (~8 nm to over 100 nm).

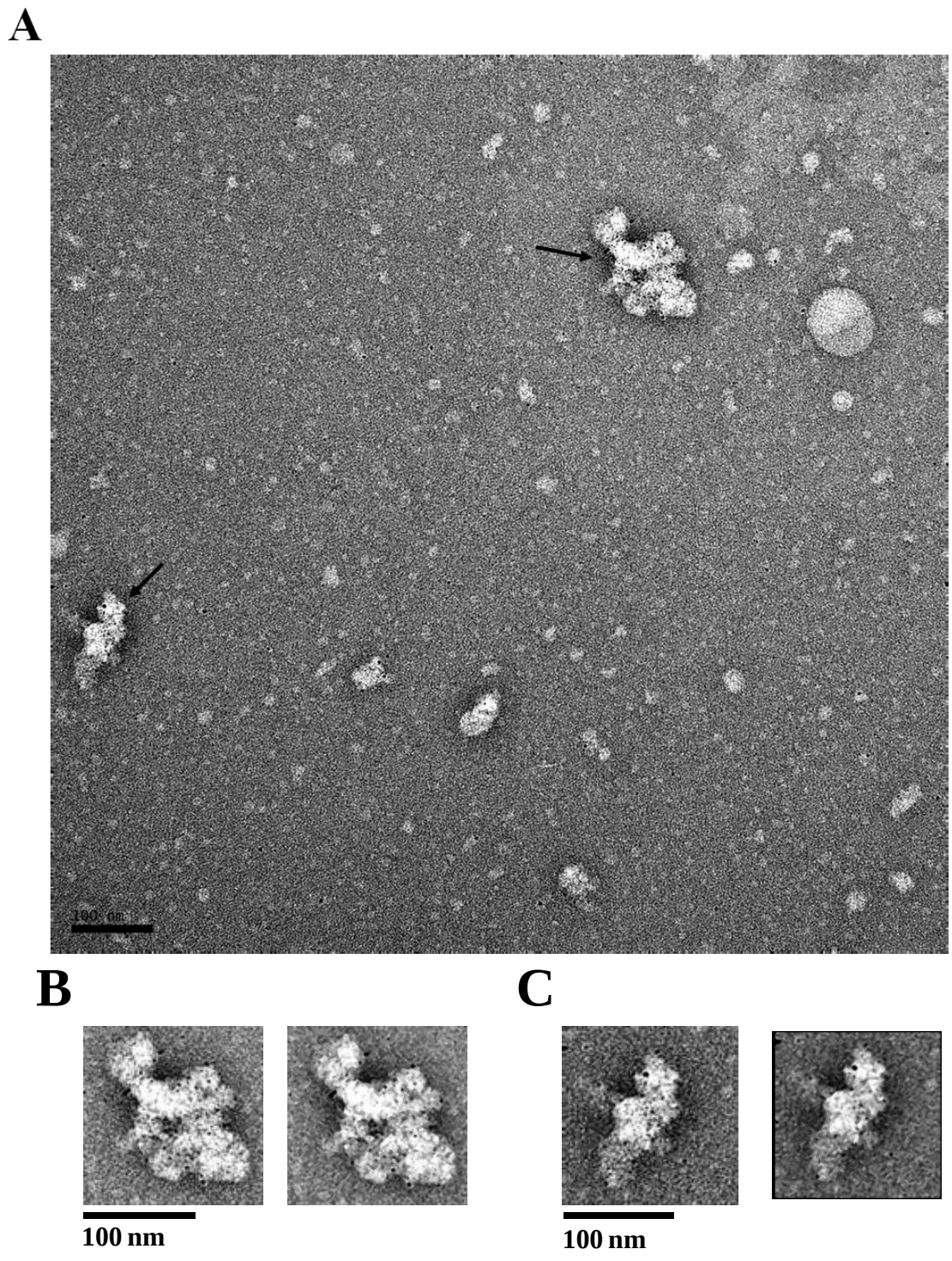
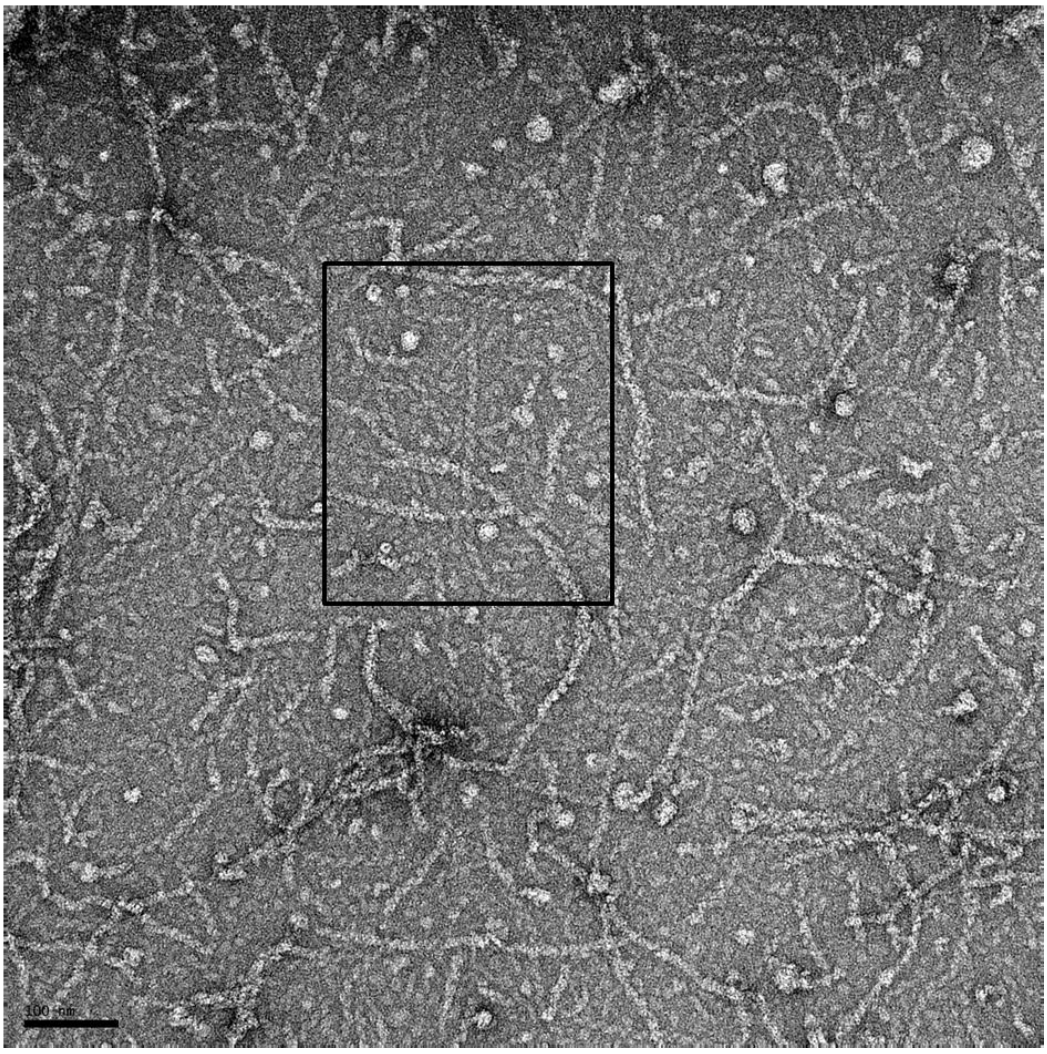



**Figure 5.5.3. Micrograph of wild-type TatAyCy-*his* void fraction + Nanogold**
*A.* The micrograph was taken at ~57,000x magnification under ~1.5 µm defocus. The grid was stained with 2% uranyl acetate. Aggregates of TatCy-*his* extensively labelled with 1.8 µm Ni-NTA-Nanogold are indicated by arrows. *B+C.* In each case the left hand-panel shows a close-up view of one of the aggregates, and the right-hand panel has been processed using the *despeckle* tool of ImageJ (ver. 1.44p) to reduce the carbon grain and enhance the Nanogold signal.

### 5.5.2. Single-particle EM of P2A mutant TatAyCy-*his* void fraction

Images taken of the corresponding void fraction for the mutant sample display a remarkably different phenotype to the wild-type as shown in Figure 5.5.4.A. An abundance of long tubular structures can be clearly seen in the undiluted sample. These are ~ 10 nm thick and range widely in length with an average spread of ~ 50-200 nm. The tubules consist of repeating units each ~ 5.5 nm long with a diagonal arrangement as highlighted in Figure 5.5.4.B and Figure 5.5.4.C. Globular particles are also present in the sample with both large, amorphous particles (up to 50 nm) and small, round and homogenous ones seen.

**A**



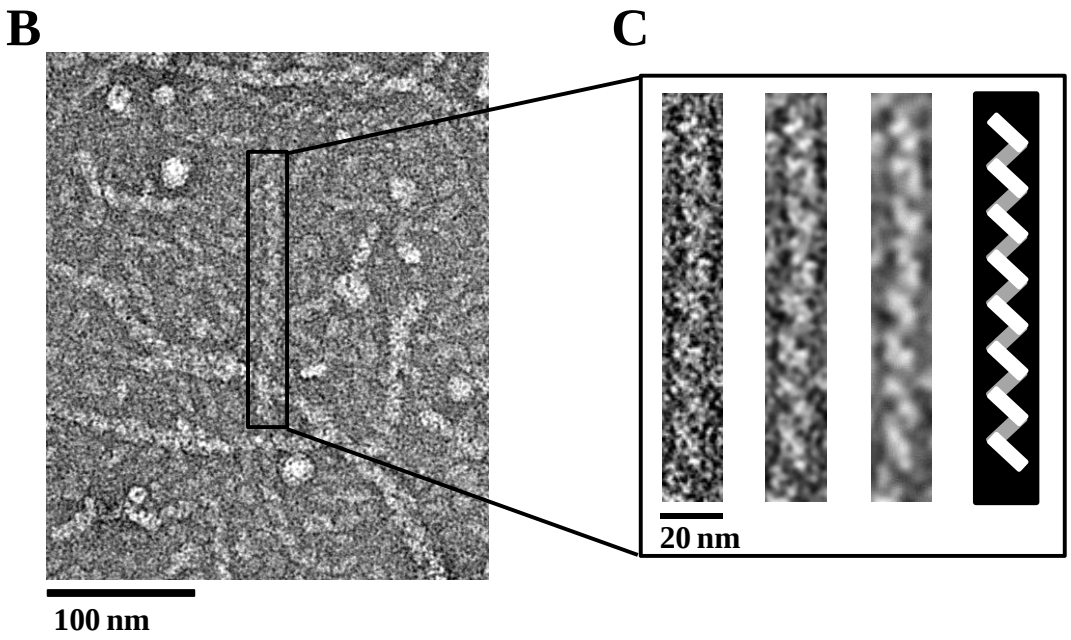



**Figure 5.5.4. Micrograph of (P2A) TatAyCy-*his* void fraction undiluted**
*A.* The micrograph was taken at ~57,000x magnification under ~1.5 µm defocus. The grid was stained with 2% uranyl acetate. *B.* Close-up of micrograph section indicated *C.* A clearly defined tubule is shown in detail. From left to right: unprocessed image, image processed with 2.0 pixel radius Gaussian filter, image processed with 4.0 pixel radius Gaussian filter, schematic representation of helical structure of tubule. Processed using the Gaussian filter tool of ImageJ (ver. 1.44p).

This diverse range of particles was not altered upon a 1:8 dilution as seen in Figure 5.5.5. Again, this suggests that the observed particle populations are not simply caused, or affected, by protein concentration. Since this fraction contains an excess of TatAy and no significant contaminants, as shown by the protein biochemistry (see Figure 5.4.7.) it can be inferred that these tubules consist primarily, if not entirely, of the mutated TatAy. Furthermore, the repeating units of these tubules fit closely to the dimensions of other TatA-type complexes observed previously (see TatAd in chapter 3, and TatE in chapter 4), suggesting that the tubules are formed by polymerisation of smaller 'ringed' TatAy complexes. Such tubule formation by *E. coli* TatA has been observed previously *in vivo* within the cytoplasm of TatA over-expressing cells (Berthelmann *et al.*, 2008). The large amorphous particles seen in this fraction closely match the morphology of those prevalent in the wild-type sample (Figure 5.5.1. to Figure 5.5.3.) suggesting that they are aggregates of, or at least contain, TatCy-*his*. The smaller ringed particles, shown more clearly in the diluted sample (Figure 5.5.5.), are also seen in the wild-type (Figure 5.5.2.) suggesting that these

constitute some form of TatAyCy-*his* complex with an unknown and potentially variable stoichiometry.

Nanogold was used again to further investigate the distribution of subunits within these particle populations. Figure 5.5.6. shows a 1:8 dilution of the mutant sample with bound 1.8 μm Ni-NTA-Nanogold® prepared in the same way as for the wild-type sample. Nanogold particles appear to be bound to the large amorphous particles and some of the smaller round particles, but not to the lengths of the tubules. Some nanogold is seen close to the extremities of tubules or to larger nodes present along their length, suggesting that residual TatCy-*his* may be present at these locations. Overall this evidence supports the theory that the tubules do not contain TatCy-*his* whilst the other particle populations do.

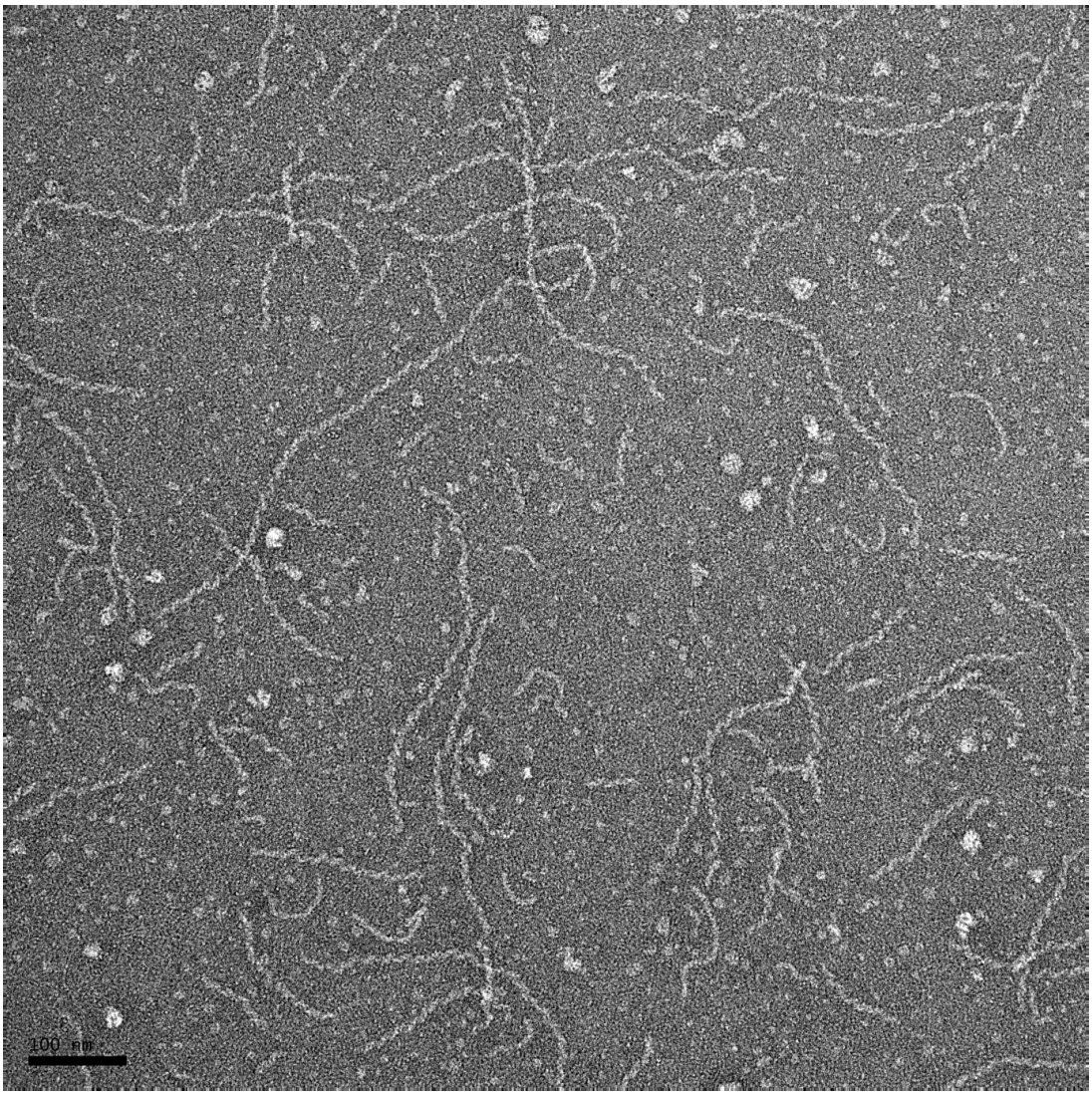



**Figure 5.5.5. Micrograph of (P2A) TatAyCy-*his* void fraction, 1:8 dilution**
The micrograph was taken at ~57,000x magnification under ~1.5 µm defocus. The grid was stained with 2% uranyl acetate.

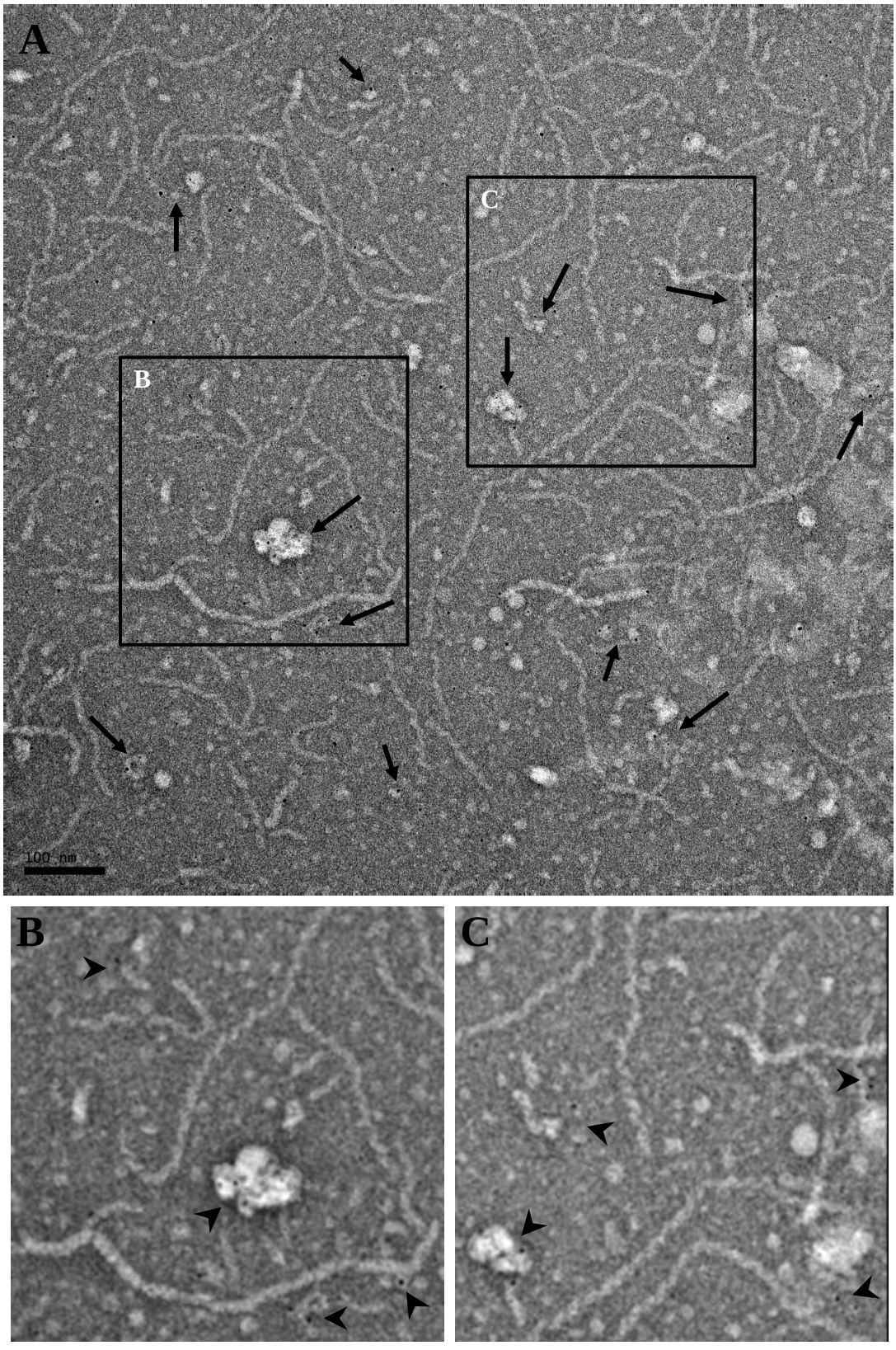
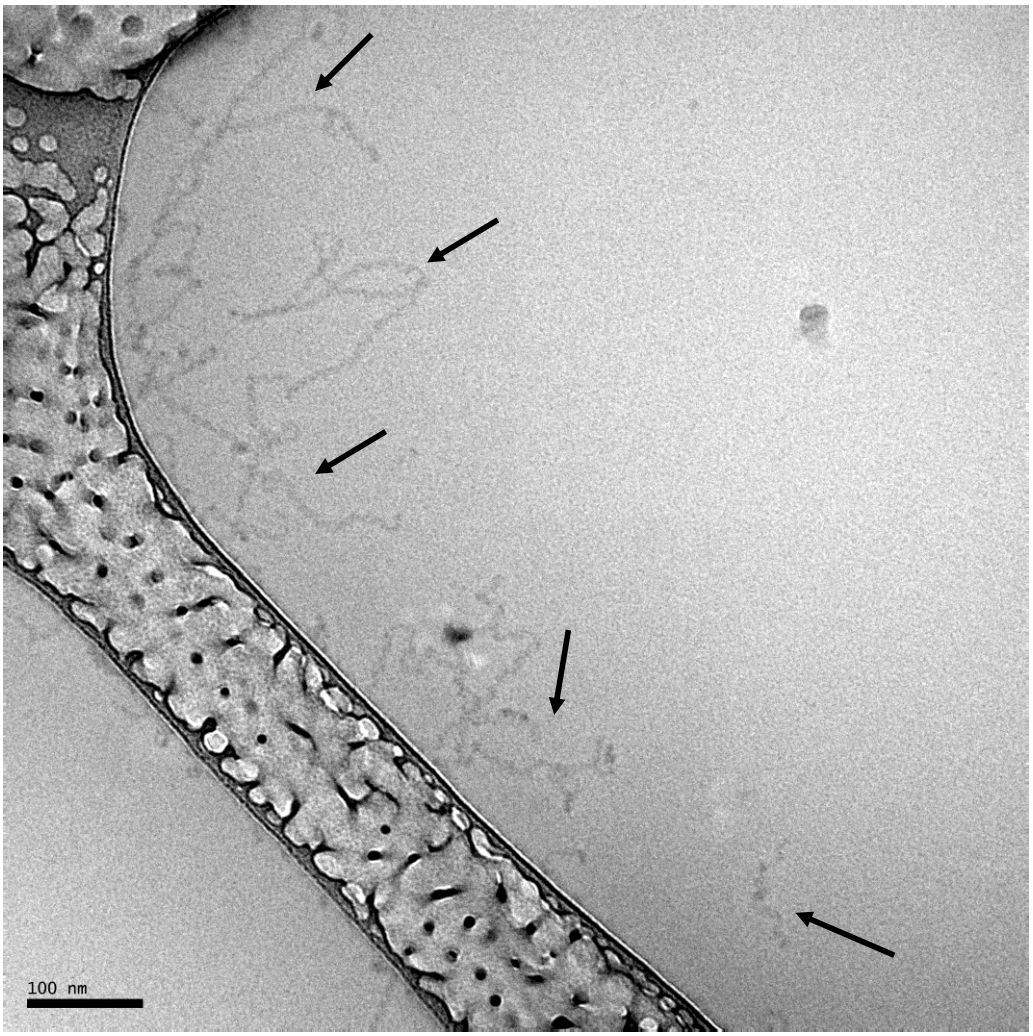



**Figure 5.5.6. Micrograph of (P2A) TatAyCy-*his* void fraction + Nanogold**
*A.* The micrograph was taken at ~57,000x magnification under ~0.5 µm defocus. The grid was stained with 2% uranyl acetate. Several clusters of Nanogold particles are indicated by arrows. *B+C.* Inset of areas indicated in *A.* 2.0 pixel Gaussian filter applied. Processed using the Gaussian filter tool of ImageJ (ver. 1.44p).

To further investigate the structure these tubules in more detail, cryo-EM grids were prepared of the undiluted sample using the method discussed previously. Over 100 Images were taken at 71565x magnification using FasTEM™ minimal dose settings (MDS) at a range of defoci. As these tubules present a larger repeating structure than smaller Tat complexes they are more suitable for imaging under the very low contrast conditions of cryo-EM. In Figure 5.5.7.A and Figure 5.5.7.B multiple tubules can be seen adopting a variety of condensed and extended conformations, the longest of these (see Figure 5.5.7.B) measuring ~ 800 nm. These data confirm that these tubules are not an artefact introduced by the harsh conditions of negative-stain; the staining and blotting procedure induces aggregation and flattening effects on the protein as well as potentially restricting the orientation of the particles on the carbon surface. Under cryo-EM conditions such effects are mitigated and the sample is imaged in a pseudo-aqueous state. These images provide high resolution data for future single particle analysis of the tubules using a filamentous particle reconstruction technique.

**A**



**B**

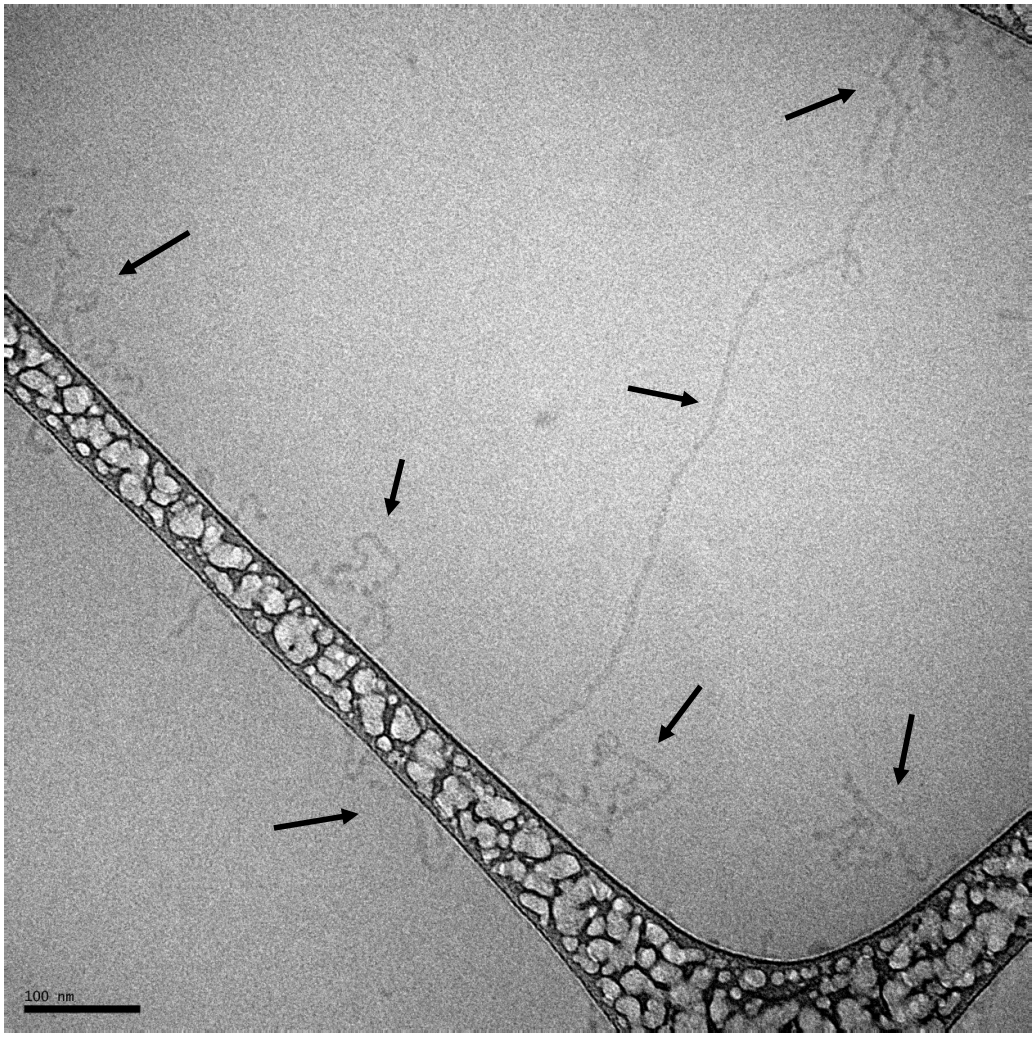



**Figure 5.5.7. Cryo-EM micrographs of (P2A) TatAyCy-*his* void fraction**
*A.+B.* Micrographs were taken at ~72,000x magnification under minimal dose settings. Samples were snap-frozen in liquid ethane to form vitreous ice. Several tubules are indicated by arrows.

### 5.5.3. Single-particle EM of wild-type TatAyCy-*his* peak fraction

Initial analysis of the peak elution fraction (GF 21, see Figure 5.4.6) reveals a very dense population of small round particles as shown in Figure 5.5.8. The dimensions of these particles are hard to determine due to the high concentration however a central pool of stain indicating a cleft or channel is visible in many of these. Some larger amorphous particles are present but these are a small minority and, as for the wild-type void fraction (see Figures 5.5.1 - 5.5.3), no tubular structures are seen.

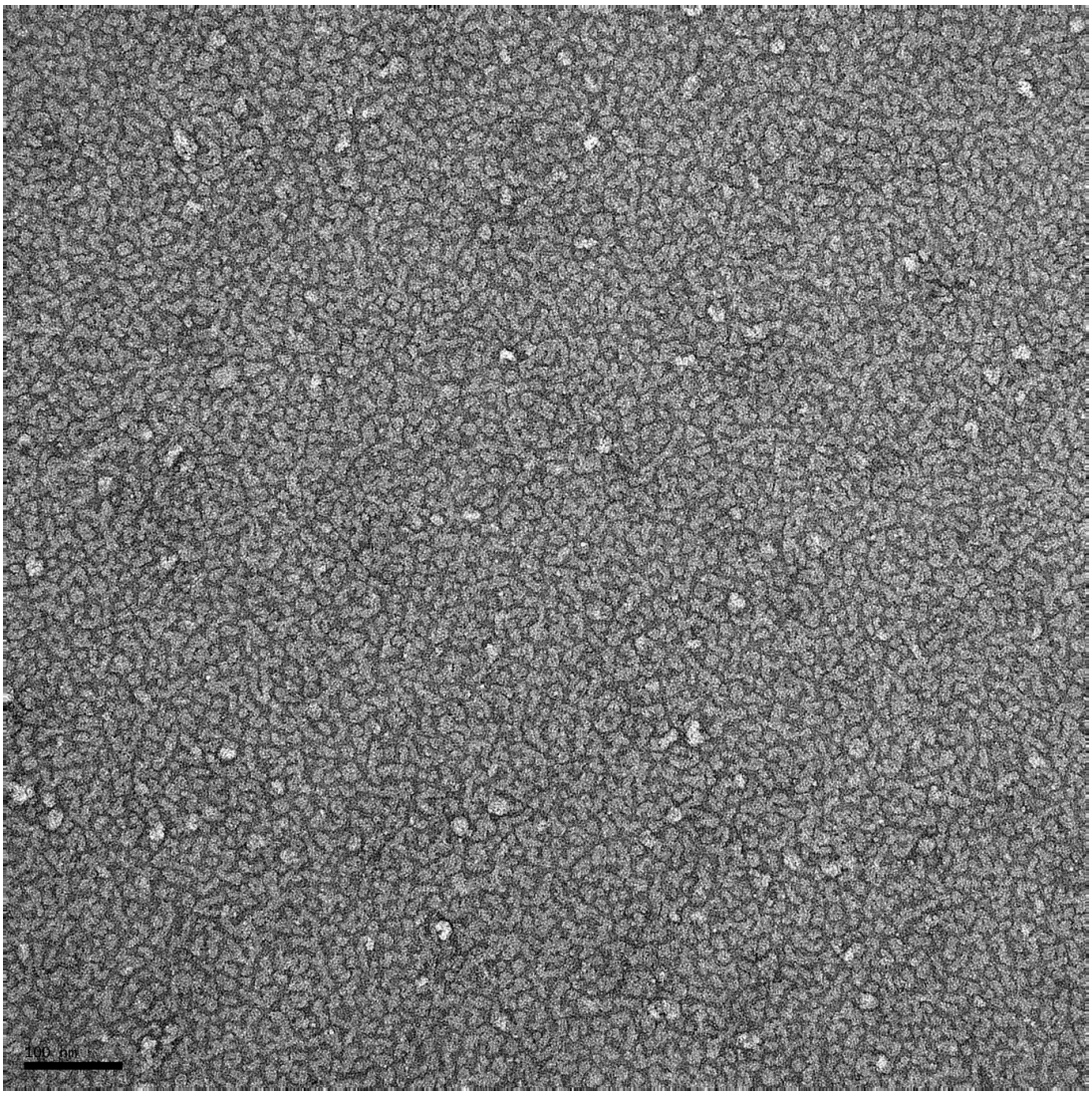



**Figure 5.5.8. Micrograph of wild-type TatAyCy-*his* peak fraction, undiluted**
The micrograph was taken at ~57,000x magnification under ~1.5 µm defocus. The grid was stained with 2% uranyl acetate.

Upon a 1:4 dilution (Figure 5.5.9.) distinct and well separated particles can be seen that are distributed within a clear size range. The larger of these particles are more amorphous and many appear to consist of multiple smaller round particles, similar to those identified for TatAyCy-*strep*. Due to the co-elution of both TatAy and TatCy-*his* within this fraction in can be assumed that these particles represent TatAyCy-*his* complexes. The vast majority of particles shown here follow the 6.0-9.5 nm size range displayed in the TatAyCy-*strep* analysis, with the larger particles ranging up to ~20 nm. None of the very large amorphous particles observed in the wild-type void fraction (50-100 nm) are seen.

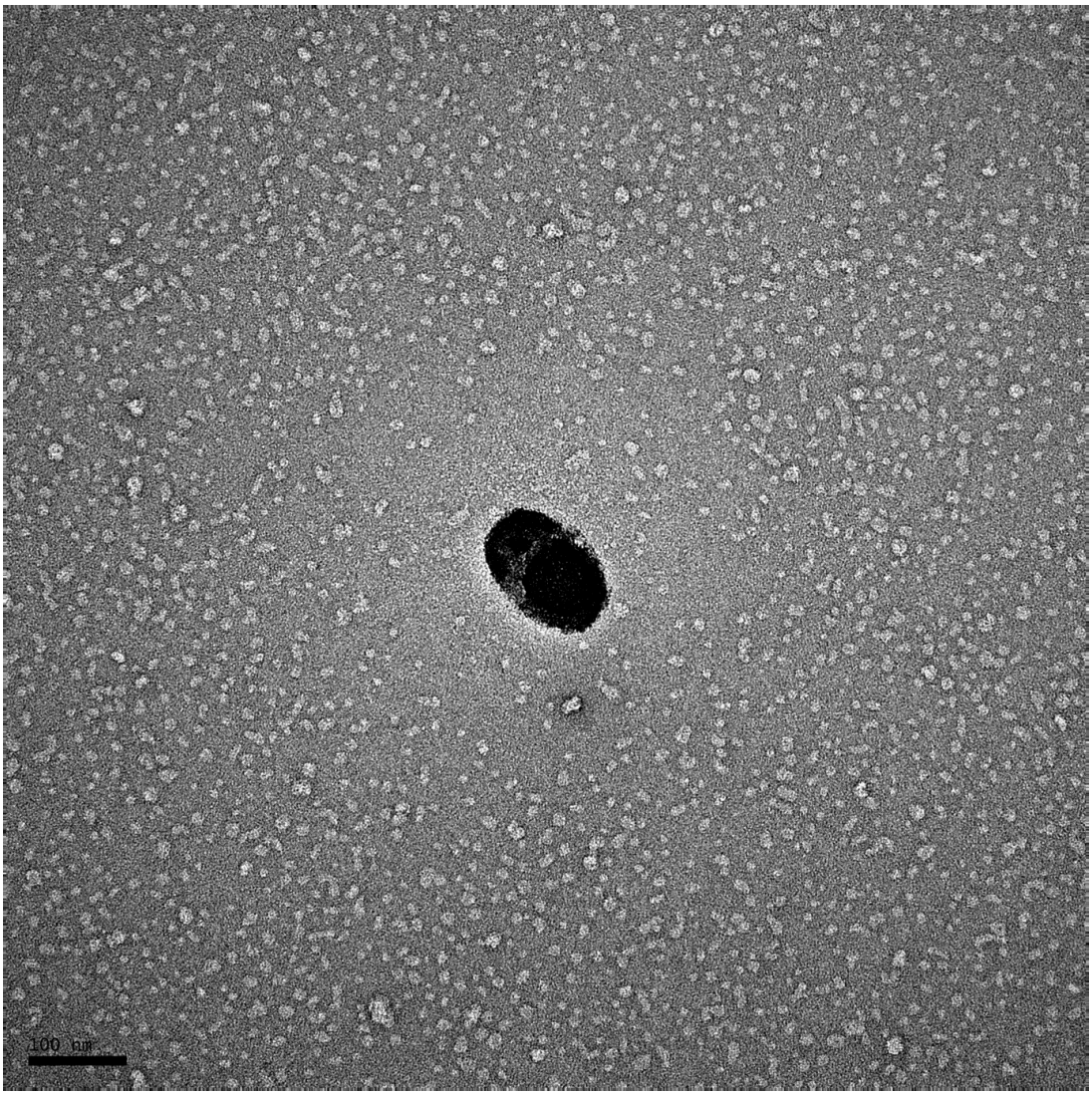



**Figure 5.5.9. Micrograph of wild-type TatAyCy-*his* peak fraction, 1:4 dilution**
The micrograph was taken at ~57,000x magnification under ~1.5 µm defocus. The grid was stained with 2% uranyl acetate. Central dark area on the grid is an artefact that provides increased contrast to the image.

Compared to the data shown for TatAd and TatE these wild-type TatAyCy particles represent the largest size range evidenced for ring-shaped complexes.

### 5.5.4. Single-particle EM of P2A mutant TatAyCy-*his* peak fraction

Both TatAy and TatCy-*his* are present in this fraction to an extent comparable to the wild-type, as shown by Western blot (see Figure 5.4.6. and Figure 5.4.7.), and this similarity extends to the particle morphology and concentration density as seen by EM. A lawn of small round particles that form the major population can be seen in Figure 5.5.10.A. (compare with Figure 5.5.9) however; clear tubules are still present as highlighted in Figure 5.5.10.B.

The distribution of these particles is clearer after a 1:4 dilution (Figure 5.5.11.), where the small particles can be clearly seen to resemble the wild-type (Figure 5.5.9.). The overall particle distribution in this fraction shows a clear contrast to those seen in the P2A mutant void fraction (compare Figure 5.5.4. to Figure 5.5.6.) and together these data point to a potential correlation between TatAy-P2A concentration, the presence of TatCy-*his*, and tubule formation.

The presence of tubules in this fraction suggests that if the localised concentration of P2A mutant TatAy is sufficiently high, polymerisation can still occur even in the presence of TatCy-*his*. As these large tubules have presumably formed after the size separation by gel filtration, these data indicate complex formation is potentially dynamic or unstable. However, the vast majority of Tat protein appears to form small homogenous particles suggesting that 'wild-type-like' TatAyCy-*his* complex formation is still possible, in turn suggesting TatCy-*his* may be able to prevent TatAy-P2A from polymerising to an extent.

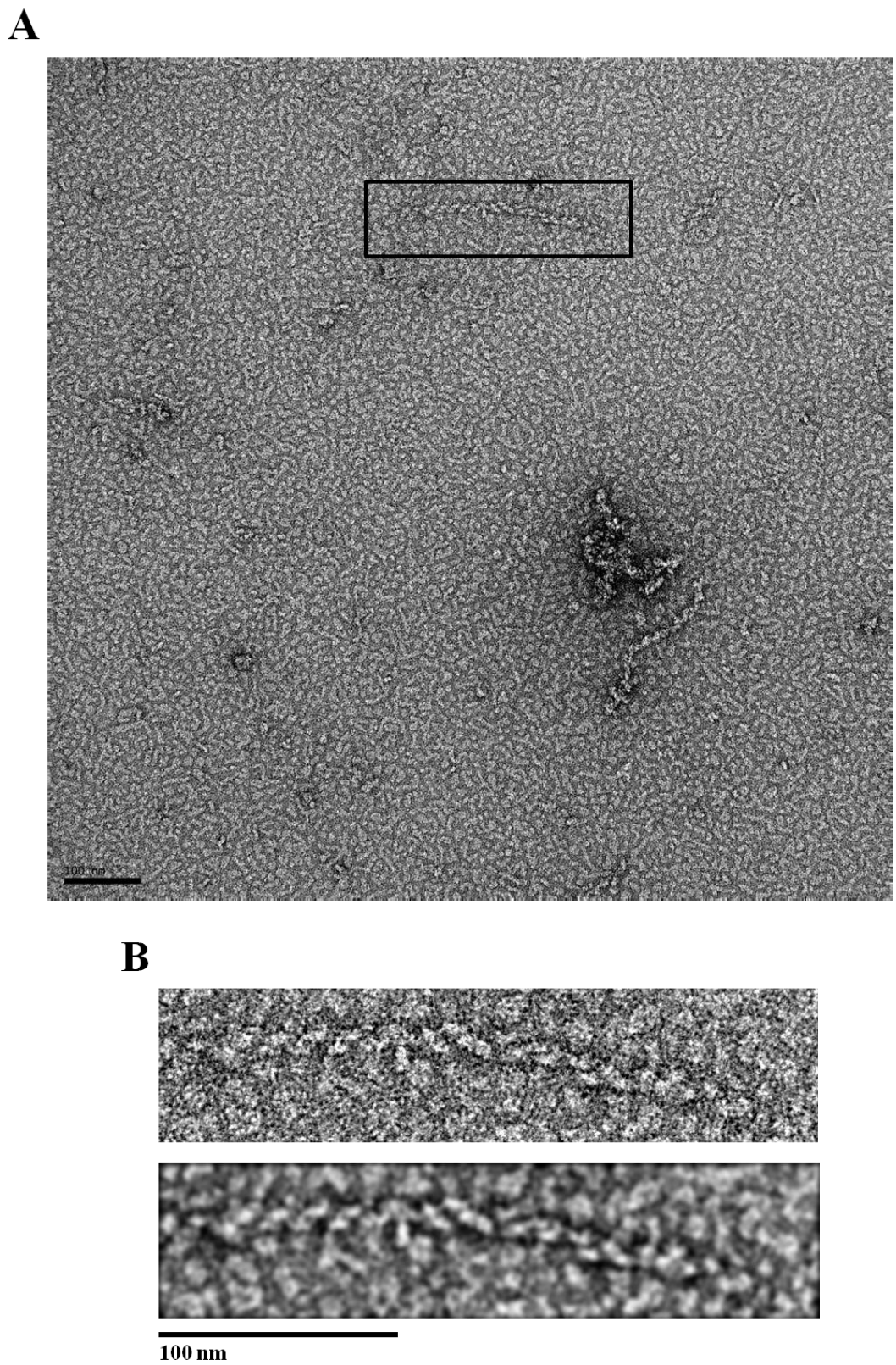



**Figure 5.5.10. Micrograph of mutant TatAyCy-*his* peak fraction, undiluted**
*A.* The micrograph was taken at ~57,000x magnification under ~1.5 µm defocus. The grid was stained with 2% uranyl acetate. *B.* Close-up of micrograph section indicated displaying a clearly defined tubule in detail. Top panel shows the unprocessed image and the bottom panel shows the same image processed with a 2.0 pixel radius Gaussian filter and despeckle tool of ImageJ (ver. 1.44p).

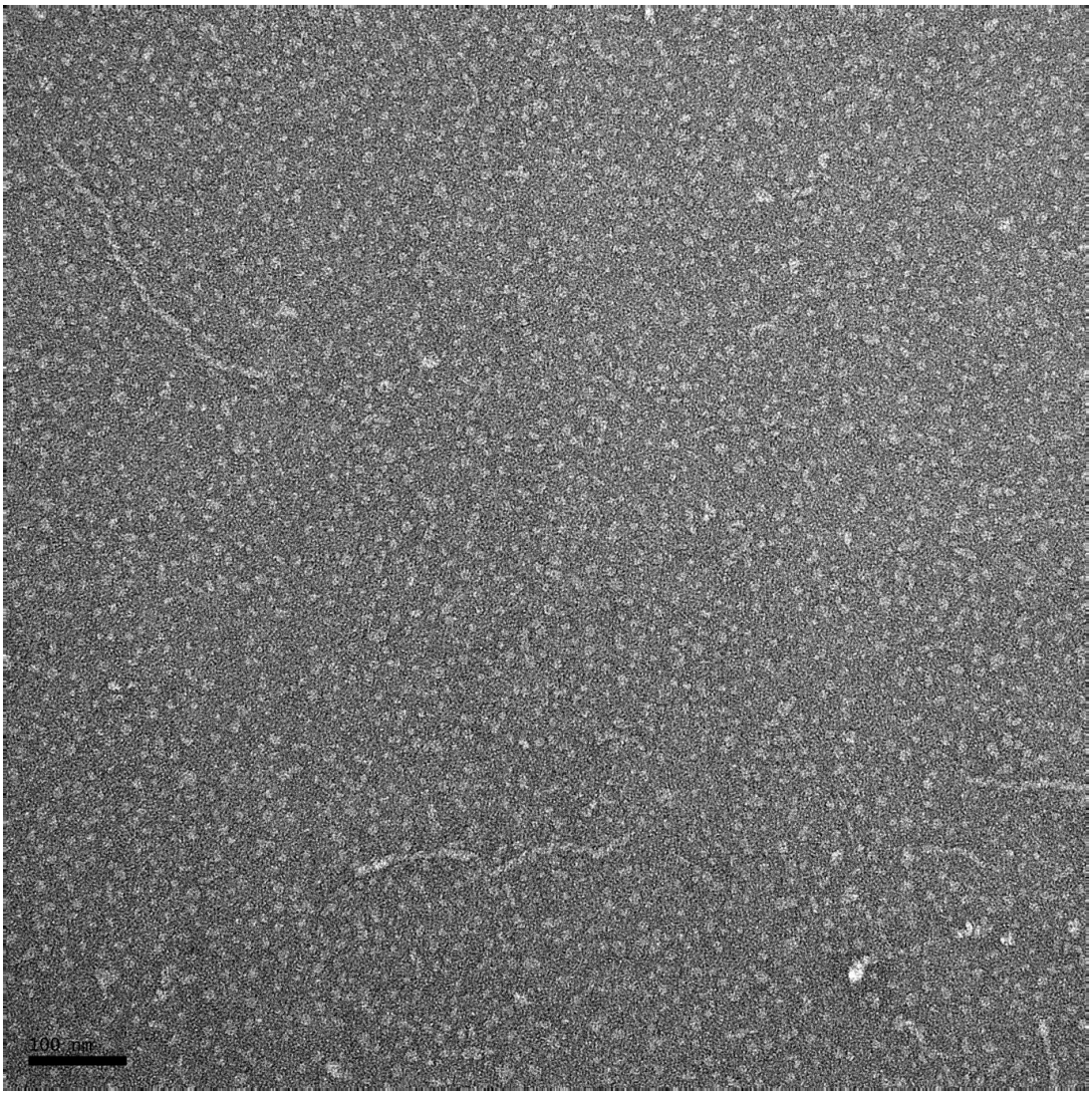



**Figure 5.5.11. Micrograph of mutant TatAyCy-*his* peak fraction, 1:4 dilution**
The micrograph was taken at ~57,000x magnification under ~1.5 µm defocus. The grid was stained with 2% uranyl acetate.

## 5.6. Discussion

In this chapter *B. subtilis* TatAyCy complexes have be subjected to an extensive structural analysis by single particle EM. These results represent the first of such an analysis on the TatAyCy pathway and present some interesting observations for the modular assembly of Tat complexes as a whole.

TatAyCy has been shown to form small, ringed particles that vary in size from ~ 6 – 9.5 nm. Along with the data presented in previous chapters for TatAd and TatE and results published for TatA (Gohlke *et al.*, 2005), this indicates that the formation of these shaped complexes is a property conserved across these three bacterial Tat pathways.

The purification of TatAyCy-*strep* showed some dissociation between TatAy and TatCy indicating that the TatAyCy complex may be structurally unstable under these conditions. However, when observed by EM no correlation between this and the size of the particle population could be seen. This suggests that the small ringed particles seen may represent a low order structurally stable unit.

The size estimations from gel filtration given here vary in some cases from those shown previously for the purification of TatAyCy (Barnett *et al.*, 2009). However, a different size exclusion column was used in this study and from the EM results provided here and in previous chapters, the relative inaccuracy of this method for such measurements has been demonstrated.

Direct evidence of complex formation between TatAy and TatCy has still yet to be presented in the form of BN-PAGE results, although the co-elution data presented here and previously is convincing. Attempts made to generate such results were unsuccessful.

From these data wild-type TatAy is shown to co-elute with TatCy-*his* across a size range of ~ 90-290 kDa, as estimated by gel filtration. No TatAy is present in higher molecular weight fractions or in the void volume. The EM investigation reveals a homogenous population of small, round particles in the peak fraction representing TatAyCy-*his* complexes, and potentially TatAy complexes, of 6-9.5 nm in diameter. Some TatCy-*his* is present in the void and forms huge aggregates as well as some small round complexes as well. No detectable TatAy is present in this fraction.

Therefore TatCy-*his* appear to have a tendency to aggregate when in excess compared to TatAy, but can still form small complexes in the presence of no/residual TatAy. This suggests TatAy is required to stabilise TatCy-*his* aggregation in TatAyCy-*his* complexes.

The P2A mutation at the extreme N-terminus of TatAy has been shown to disrupt the translocation of YwbN in both *B. subtilis* and *E. coli* translocation assay systems (van der Ploeg *et al.*, 2011). In this study no significant disruption to the structure of the complexes was detected, however in this case the proteins were expressed in an *E. coli* system and analysis was performed via gel filtration only.

In the data presented here the presence of the P2A mutation was shown to encourage the majority of TatAy to polymerise and form large, distinct tubules, the segmented structure of which suggests a modular arrangement. The dimensions of these segments, along with the presence of small ringed particles in the same fraction, indicate that the normal protein-protein interactions between TatAy subunits have not been overly-disrupted by this point-mutation. Rather, it seems that new interactions between TatAy complexes have been strongly promoted. Following the accepted 'L-shaped' orientation of TatA-like subunits within multimeric complexes suggests that the N-terminus of one complex is now interacting with the hinge region or amphipathic helix of the next, resulting in a continuous spiral formation. A model representing this hypothetical arrangement is shown in detail in Figure 5.6.1.

Tubule formation by Tat components has been observed previously for the *E. coli* system *in vivo* (Berthelmann *et al.*, 2008). In this largely over-looked publication tubule formation was shown clearly when over-expressing wild-type TatABC. As this tubule formation required the presence of TatC it would be interesting to see if such a requirement was true of *B. subtilis* TatCy. Upon immunogold labelling against all three *E. coli* Tat subunits only TatA could be detected. TatAy tubules shown here did form in the presence of TatCy and Nanogold labelling could not detect TatCy-*his* in the body of the tubules. Interestingly in the same study extensions were seen to interconnect various tubules to form an array of connected rings when viewed from a top-down perspective. Such an arrangement bears a similarity to the connections seen between the tetramer units shown here in the TatAyCy-*strep* data, and

furthermore Nanogold labelling did detect TatCy in nodules or strands on the outside of the TatAy tubules.

The authors of this report suggested that the single ring structures observed in other Tat studies represent a detergent-induced disruption of normal Tat complexes. Following this line of enquiry it may be that the P2A mutation acts to preserve an *in vivo* structure that is normally destroyed upon cell disruption and detergent solubilisation.

They also indicate in this study that C-terminal fusions of fluorescent proteins to TatA inhibit the tubal formation, potentially due to steric hindrance, whereas if fluorescent tagged Tat is co-expressed with native *tatABC* genes fluorescent tubules are visible. This may explain why tubules were not detected in another *in vivo* study where single molecule fluorescence was used (Leake *et al.*, 2008). Interestingly in this study TatA complexes could not be detected in the absence of TatBC raising further questions about the validity of the pore-fitting model (Gohlke *et al.*, 2005).

The data presented in previous chapters has clearly shown the major impact that detergent concentration has upon the observed particle population, it therefore cannot be assumed that simply because a sample appears stable under given conditions it is biologically significant. In this way the aggregates indicated for TatAd (chapter 3) upon reduction in DDM concentration to below the CMC may in fact represent disorganised polymerisation rather than completely irrelevant aggregation.

In summary the data presented here indicate that small ring shaped complexes may be the building blocks for higher order Tat structures and that the N-terminus of TatAy has an important role in protein-protein interactions within these complexes.

It would be very interesting to visualise the other TatAy mutants that have been characterised (van der Ploeg *et al.*, 2011) to affect translocation to determine whether they show a similarly drastic phenotype.

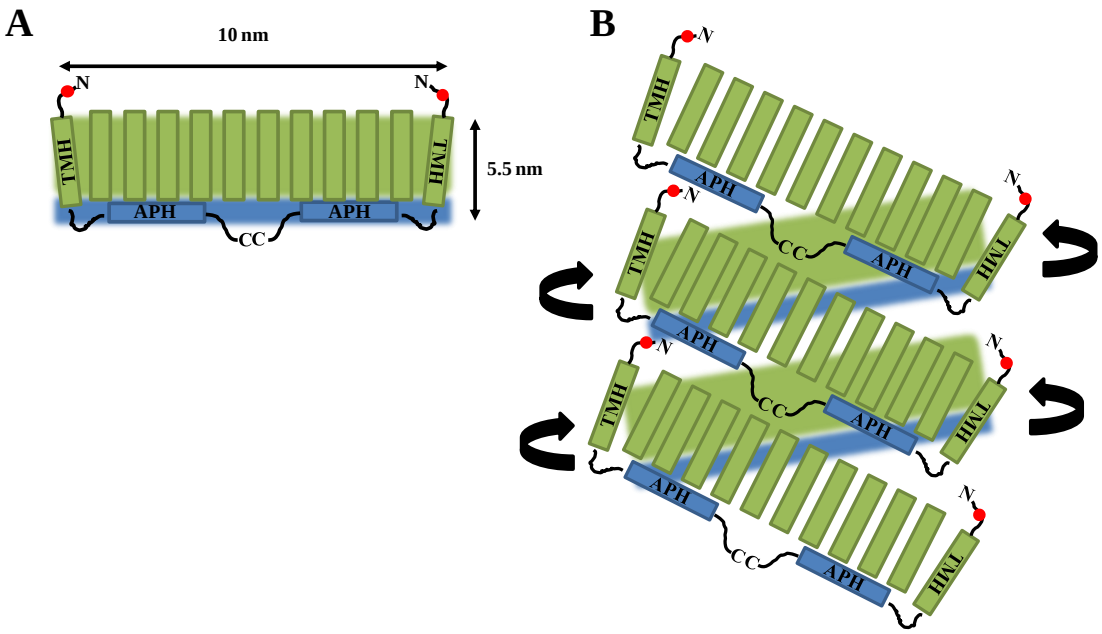



**Figure 5.6.1. Schematic model of proposed TatAy complex arrangement in the formation of observed tubules**

*A.* Side view (in the plane of a hypothetical membrane) of TatAy ring complex based on the accepted assembly of TatAy monomers. The ring is formed by multiple transmembrane helices (TMH, in green) and the proposed 'lid' is formed by the amphipathic helices (APH, in blue). The site of the P2A point mutation is shown near the N-termini of the outermost monomers (red dots). *B.* Proposed arrangement of these ringed units to form the observed continuous helical assemblies. The point mutation near the N-termini of one TatAy monomer promotes a protein-protein interaction with the hinge region or APH of another. As enough TatAy monomers assemble to form a ring of ~ 10 nm width, the mutation promotes the continuation of subunit polymerisation. The newly promoted protein-protein interaction induces a pitch sufficient to allow the monomers within each turn to interact with the opposite face of the monomers within the next turn. The directionality of the continuous helix is indicated by the curved arrows. The bands of colour in the background represent the continuation of the helix on the opposite side.

# *Chapter 6*

# *Final discussion*

## 6.1. Final discussion

Over the years the majority of studies into the bacterial Tat pathway have focused on the *E. coli* system; as is typical of Gram-negative bacteria, this consists of three essential components: TatA, TatB and TatC (Bogsch *et al.*, 1998; Sargent *et al.*, 1998; Sargent *et al.*, 1999). More recently *Bacillus subtilis* has emerged as the Gram-positive model organism for bacterial Tat studies, where two minimal Tat pathways, consisting of only type-A and type-C subunits, have been identified to date (Barnett *et al.*, 2008; Yen *et al.*, 2002).

In the currently accepted model of Tat-mediated translocation, the substrate initially interacts with a Tat(A)BC substrate recognition complex (Alami *et al.*, 2003; Alami *et al.*, 2002; Cline and Mori, 2001; McDevitt *et al.*, 2006) which in turn triggers the transient coalescence of a separate TatA complex, or complexes (Mori, Hiroki and Cline, 2002), to form a size-fitting pore through which the substrate can traverse the membrane bilayer (Gohlke *et al.*, 2005; Oates *et al.*, 2005).

In this model much mechanistic relevance has been placed on the observed size range present in purified *E. coli* TatA complexes (Gohlke *et al.*, 2005; Oates *et al.*, 2005). However, biochemical data from a number of more recent studies has indicated that such a distinct variation in complex size is not a feature common across all Tat complexes (Baglieri *et al.*, 2012; Barnett *et al.*, 2008; van der Ploeg *et al.*, 2011).

In this investigation single-particle EM has been used as a basis for structural studies into a range of bacterial Tat complexes from both Gram-negative *E. coli* and Gram-positive *B. subtilis*. Direct 2D and 3D structural models of a range of bacterial Tat complexes have been presented for the first time; this includes the first structural investigation into complex formation by translationally deficient Tat mutants. The work presented in this thesis largely supports previous biochemical observations about *B. subtilis* Tat complexes (Barnett *et al.*, 2008; Barnett *et al.*, 2009); the main conclusion being that they appear relatively homogeneous in size compared to those of *E. coli.*

Membrane integral TatAd complexes, expressed from an *E. coli ΔtatABCDE* strain, were isolated and purified for direct visualisation by TEM. TatAd complexes were

shown to consist of small ring-shaped structures of 7.5-9.0 nm in diameter, with a central cavity or pore of 2.5-3.0 nm. The gross morphology of these complexes (ring-shaped) was consistent with those observed previously for other Tat complexes using EM techniques (Gohlke *et al.*, 2005; Oates *et al.*, 2003; Sargent, F *et al.*, 2001; Westermann *et al.*, 2006). Using an established, size differentiating, statistical classification method (White *et al.*, 2004) TatAd complexes were divided into 4 size classes. However, TatAd complexes did not form the large number of differently sized structures that were seen previously for *E. coli* TatA (Gohlke *et al.*, 2005). TatA complexes were identified in a size range of 9-13 nm with a corresponding variance in pore diameter of 3-7 nm. The authors of this study indicated that the largest identified class of TatA complexes, with a 7 nm pore, was sufficient to translocate the large, 90 kDa Tat substrate, TorA.

Interestingly, the smallest class of TatAd complexes presented here showed potential extensions from the globular densities that form the ring structure. To account for this, a model in which the TatAd subunit undergoes a conformational change was proposed. This involved the bending of the hinge region to place the amphipathic helix in juxtaposition with the transmembrane helix, resulting in a 1-2 nm extension from the plane of the ring. This model was based on a similar conformational shift that has been proposed previously to account for an observed dual topology for *E. coli* TatA complexes (Gouffi *et al.*, 2004). These extensions were not observed in the larger TatAd complexes where the central cavity was clearly occluded from one side. This in turn suggested that TatAd complexes were present in a mixture of 'open' and 'closed' states.

Importantly no TatAd complexes were identified here that present an internal cavity large enough to accommodate TorA. In fact, the average TatAd complex measured only 8 nm, even smaller than the smallest class of TatA, previously identified by Gohlke et al. (2005) at 9 nm. As translocation assays have shown that TatAd can substitute for TatA to transport TorA (Barnett *et al.*, 2008), this calls the proposed relevance of the TatA size range into question; a line of enquiry further supported by observations involving other Tat complexes. It has been noted that the distinctive ladder of BN-PAGE bands associated with TatA (Oates *et al.*, 2005) can in fact be effectively replicated using TatB, when over-expressed in the presence or absence of TatC (Behrendt *et al.*, 2007); in this study the authors proposed that the ladders of

TatB (and TatA) may simply represent the disassembly of a larger structure. The observed ladder of TatB ranged from ~ 100 – 900 kDa (Behrendt *et al.*, 2007), similar to the TatA ladder of ~ 100-700 kDa (Oates *et al.*, 2005). Furthermore, in the model plant *Arabidopsis thaliana* TatA was found to form complexes below 200 kDa by BN-PAGE and was detectable only in minor amounts compared to TatB and TatC (Jakob *et al.*, 2009). Additionally, *in vivo* fluorescence studies on TatA stoichiometry were able to identify signs of complex formation in the membrane, but only in the presence of TatB and TatC (Leake *et al.*, 2008). This body of evidence suggests that the significance of the 'size-fitting pore' model may require further consideration.

There also exists strong evidence to support the bifunctional nature of TatAd: its ability to assume both TatA and TatB roles (Barnett *et al.*, 2008; Jongbloed *et al.*, 2006). Despite this functional overlap, TatAd was shown to form discrete complexes by gel filtration (Barnett *et al.*, 2008; Barnett *et al.*, 2009) in contrast to the diffuse profile of TatA complexes (Oates *et al.*, 2005). The structural data presented here confirms these findings, with direct EM visualisation of small, discrete TatAd complexes, that contrast with previous results for *E. coli* TatA (Gohlke *et al.*, 2005).

Next the TatA paralogue, TatE, from *E.coli* was investigated using similar EM methods. A single-particle EM analysis was performed on TatE purified from an *E. coli ΔtatAE* strain. TatE was found to form small ring-shaped complexes of less than 8 nm in diameter; these were separated into 2 size classes measuring ~ 6 nm and ~ 7.5 nm respectively, with a central cavity of 2.5 nm occluded on one side. No evidence was found of TatE complexes large enough to accommodate folded Tat substrates using either biochemical or EM classification methods. 3D electron density maps were generated of these structures to a final resolution of 29 Å.

As TatE can complement a *ΔtatA* mutant (Sargent *et al.*, 1999) and can translocate several Tat substrates (Baglieri *et al.*, 2012) these data appear inconsistent with the current 'size-fitting pore' model for TatA-type complexes (Gohlke *et al.*, 2005). The observation of a significant population of multi-ringed complexes within the TatE particle-set suggested that these may represent modular assembly intermediates of a larger super-complex. A sub-population of short, rod-shaped structures was also identified, that corresponded with the dimensions of the multi-ringed structures. However, although a crude 3D reconstruction revealed the presence of an internal

cavity, no signs of a subdivision into 2 openings could be determined. From this data it was not possible to ascertain the orientational relationship of this particle class relative to the single-ringed or multi-ringed classes. Future investigation using a range of different staining and gridding techniques to manipulate the orientational preference of the TatE complexes could be used to probe this further. Additionally, cryo-preparation could be used to investigate whether the rod shaped particles represent a flattening artefact introduced during negative staining.

In chapter 5 a structural investigation was conducted into the *B. subtilis* TatAyCy complex, the constituently expressed type-AC Tat system. TatAyCy was over-expressed in *B. subtilis* and purified to homogeneity before EM analysis. This study revealed that TatAyCy, like other Tat complexes, consists of small ring-shaped structures; these varied in size from 6.0-9.5 nm. By observing a range of gel filtration elution fractions under EM, the relationship between complex size and relative subunit stoichiometry was investigated. Excess TatCy was seen to aggregate in the absence of TatAy, suggesting that its presence is required to stabilise TatCy in TatAyCy complexes. The presence of TatCy in these aggregates was confirmed using Nanogold labelling. A range of multi-ringed structures were also identified in this study, including a small number of potential tetrameric Tat complexes of ~ 17 nm diameter with a central cavity of ~ 5.5 nm. This represented the largest 'pore' observed in any of the Tat complexes investigated during this work. Interestingly, these tetramers were only identified in samples where both the TatAy and TatCy subunits were present. These observations would be concurrent with the theory of modular super-complex formation by these smaller Tat complexes. However, although striking, these tetrameric structures were seen in very low abundance compared to the single ringed-structures, suggesting that they may be unstable under these experimental conditions, or possibly reflecting their transient coalescence *in vivo*. Either way, until such 'super-complexes' can be isolated in greater numbers this observation cannot be deemed truly significant.

Taken together, the structures produced during this investigation do not agree with the currently accepted 'size-fitting pore' model of Tat mediated translocation. These data support better the range of alternative models (Brüser and Sanders, 2003; Cline and McCaffery, 2007; Dabney-Smith *et al.*, 2006), which predict localised concentrations of Tat proteins act to destabilise the membrane bilayer. Such models

focus on the transient-coalescence of TatA-type subunits and fit well with the observed clustering of multiple, small, ringed complexes seen throughout the work shown here. However, the implications of these models on future structural work are intimidating to say the least. In the absence of a permanent channel structure, even more emphasis is placed on the ability to extract and purify a stalled Tat translocon with associated substrate.

The work presented here represents a significant addition to the structural profiling of bacterial Tat complexes as a whole. Nevertheless, the impact of detergent extraction and the contribution of detergent micelles to these structures remains a continuing problem to the field. The impact of detergent concentration on the morphology of these complexes has been shown to be highly significant if reduced below the critical micelle concentration (CMC) (see Chapter 3 + 4). However, by maintaining the concentration close to the CMC it appears that the detergent contribution can be reduced to minimal levels.

In order to circumnavigate the detergent issue, attempts were made to employ SMALP [poly(styrene-co-maleic acid)] Nanodiscs (Knowles *et al.*, 2009) as a means of membrane protein extraction. These have been used successfully in the past to extract membrane proteins and result in an environment more reminiscent of the membrane bilayer. However, at present it appears that the 10 nm maximal width of these discs is somewhat restrictive for Tat studies. Although SMALPs were used to successfully extract *B. subtilis* TatAd, downstream purification and imaging via EM proved difficult. Further work with these nanoparticles remains an interesting and potentially very fruitful avenue for the future.

Finally, a mutant TatAyCy complex was analysed and compared to the wild-type using the same biochemical and EM techniques employed previously. In the TatAy P2A mutant a single point mutation introduced an amino-acid substitution at the extreme N-terminus of the protein. EM analysis revealed this mutation to introduce a drastic phenotypic change, resulting in the formation of long tubular structures. These had an invariable thickness of ~ 10 nm but varied widely in length, with an average spread of 50-200 nm but reaching up to 800 nm long. The tubules consisted of repeating modular units of ~ 5.5 nm in length, arranged diagonally with respect to each-other. Nanogold labelling indirectly revealed that these tubules consisted

primarily of TatAd. To explain these tubules, a hypothesis involving the promotion of protein-protein interactions between the mutated N-terminus of TatAy and the hinge-region of an adjacent subunit was proposed.

The importance of the extreme N-terminal region of TatA-type proteins has been demonstrated previously through a number of mutagenesis studies (Barnett *et al.*, 2011; Blaudeck *et al.*, 2005; van der Ploeg *et al.*, 2011). This region has also been linked to TatB-type functionality, as substitutions in *E. coli* TatA allow for TorA export in the absence of TatB (Blaudeck *et al.*, 2005) and abolish the bifunctionality of *B. subtilis* TatAd (Barnett *et al.*, 2011) and TatAy (van der Ploeg *et al.*, 2011). More recently the N-terminal region of *E. coli* TatA has been implicated in an interaction with PspA, a membrane stress response protein, with which it was shown to co-purify (Mehner *et al.*, 2012). Together, these data, along with that presented here, strongly indicate that the N-terminus of TatA-type proteins is involved in regulating protein-protein interactions during complex formation.

The visualisation of other translocation deficient Tat mutants remains an important avenue for future work. The observation of similarly drastic phenotypes, in combination with subunit specific labelling, could reveal much about the dynamics of complex assembly, stoichiometry and even the Tat translocation mechanism.

# *Chapter 7*

# *References*

**Alami, M., Lüke, I., Deitermann, S., Eisner, G., Koch, H.-G., Brunner, J. and Müller, M.** (2003). Differential Interactions between a Twin-Arginine Signal Peptide and Its Translocase in Escherichia coli. *Molecular Cell*, **12**, 937-946.

**Alami, M., Trescher, D., Wu, L.-F. and Muller, M.** (2002). Separate Analysis of Twin-arginine Translocation (Tat)-specific Membrane Binding and Translocation in Escherichia coli. *Journal of Biological Chemistry*, **277**, 20499-20503.

**Aldridge, C., Spence, E., Kirkilionis, M. A., Frigerio, L. and Robinson, C.** (2008). Tat-dependent targeting of Rieske iron-sulphur proteins to both the plasma and thylakoid membranes in the cyanobacterium Synechocystis PCC6803. *Molecular Microbiology*, **70**.

**Allen, S. C. H., Barrett, C. M. L., Ray, N. and Robinson, C.** (2002). Essential Cytoplasmic Domains in the Escherichia coli TatC Protein. *Journal of Biological Chemistry*, **277**, 10362-10366.

**Bachmann, J., Bauer, B., Zwicker, K., Ludwig, B. and Anderka, O.** (2006). The Rieske protein from Paracoccus denitrificans is inserted into the cytoplasmic membrane by the twin-arginine translocase. *FEBS J*, **273**, 4817-30.

**Backert, S. and Meyer, T. F.** (2006). Type IV secretion systems and their effectors in bacterial pathogenesis. *Curr Opin Microbiol*, **9**, 207-17.

**Bageshwar, U. K. and Musser, S. M.** (2007). Two electrical potential-dependent steps are required for transport by the Escherichia coli Tat machinery. *J Cell Biol*, **179**, 87-99.

**Baglieri, J., Beck, D., Vasisht, N., Smith, C. J. and Robinson, C.** (2012). Structure of TatA Paralog, TatE, Suggests a Structurally Homogeneous Form of Tat Protein Translocase That Transports Folded Proteins of Differing Diameter. *J. Biol. Chem.*, **287**, 7335-7344.

**Barnett, J. P., Eijlander, R. T., Kuipers, O. P. and Robinson, C.** (2008). A Minimal Tat System from a Gram-positive Organism: A BIFUNCTIONAL TatA SUBUNIT PARTICIPATES IN DISCRETE TatAC AND TatA COMPLEXES. *J. Biol. Chem.*, **283**, 2534-2542.

**Barnett, J. P., Lawrence, J., Mendel, S. and Robinson, C.** (2011). Expression of the bifunctional Bacillus subtilis TatAd protein in Escherichia coli reveals distinct TatA/B-family and TatB-specific domains. *Arch Microbiol*, **193**, 583-94.

**Barnett, J. P., van der Ploeg, R., Eijlander, R. T., Nenninger, A., Mendel, S., Rozeboom, R., Kuipers, O. P., van Dijl, J. M. and Robinson, C.** (2009). The twin-arginine translocation (Tat) systems from Bacillus subtilis display a conserved mode of complex organization and similar substrate recognition requirements. *FEBS Journal*, **276**, 232-243.

**Barrett, C. M. L., Freudl, R. and Robinson, C.** (2007). Twin Arginine Translocation (Tat)-dependent Export in the Apparent Absence of TatABC or TatA Complexes Using Modified Escherichia coli TatA Subunits That Substitute for TatB. *Journal of Biological Chemistry*, **282**, 36206-36213.

**Barrett, C. M. L., Mangels, D. and Robinson, C.** (2005). Mutations in Subunits of theEscherichia coli Twin-arginine Translocase Block Function via Differing Effects on Translocation Activity or Tat Complex Structure. *Journal of Molecular Biology*, **347**, 453-463.

**Baxter, W. T., Leith, A. and Frank, J.** (2007). SPIRE: The SPIDER Reconstruction Engine. *Journal of Structural Biology*, **157**, 56-63.

**Behrendt, J., Lindenstrauß, U. and Brüser, T.** (2007). The TatBC complex formation suppresses a modular TatB-multimerization in Escherichia coli. *FEBS Letters*, **581**, 4085-4090.

**Behrendt, J., Standar, K., Lindenstrauß, U. and Brüser, T.** (2004). Topological studies on the twin-arginine translocase component TatC. *FEMS Microbiology Letters*, **234**, 303-308.

**Beloin, C., Valle, J., Latour-Lambert, P., Faure, P., Kzreminski, M., Balestrino, D., Haagensen, J. A., Molin, S., Prensier, G., Arbeille, B. and Ghigo, J. M.** (2004). Global impact of mature biofilm lifestyle on Escherichia coli K-12 gene expression. *Mol Microbiol*, **51**, 659-74.

**Berg, B. v. d., Clemons, W. M., Collinson, I., Modis, Y., Hartmann, E., Harrison, S. C. and Rapoport, T. A.** (2004). X-ray structure of a protein-conducting channel. *Nature*, **427**, 36-44.

**Berks, B. C.** (1996). A common export pathway for proteins binding complex redox cofactors? *Mol Microbiol*, **22**, 393-404.

**Berks, B. C., Sargent, F. and Palmer, T.** (2000). The Tat protein export pathway. *Molecular Microbiology*, **35**, 260-274.

**Bernhardt, T. G. and de Boer, P. A.** (2003). The Escherichia coli amidase AmiC is a periplasmic septal ring component exported via the twin-arginine transport pathway. *Mol Microbiol*, **48**, 1171-82.

**Berthelmann, F., Mehner, D., Richter, S., Lindenstrauss, U., Lunsdorf, H., Hause, G. and Bruser, T.** (2008). Recombinant Expression of tatABC and tatAC Results in the Formation of Interacting Cytoplasmic TatA Tubes in Escherichia coli. *J. Biol. Chem.*, **283**, 25281-25289.

**Blaudeck, N., Kreutzenbeck, P., Freudl, R. and Sprenger, G. A.** (2003). Genetic analysis of pathway specificity during posttranslational protein translocation across the Escherichia coli plasma membrane. *J Bacteriol*, **185**, 2811-9.

**Blaudeck, N., Kreutzenbeck, P., Muller, M., Sprenger, G. A. and Freudl, R.** (2005). Isolation and Characterization of Bifunctional Escherichia coli TatA Mutant Proteins That Allow Efficient Tat-dependent Protein Translocation in the Absence of TatB. *Journal of Biological Chemistry*, **280**, 3426-3432.

**Bogsch, E., Brink, S. and Robinson, C.** (1997). Pathway specificity for a delta pH-dependent precursor thylakoid lumen protein is governed by a 'Sec-avoidance' motif in the transfer peptide and a 'Sec-incompatible' mature protein. *EMBO J*, **16**, 3851-9.

**Bogsch, E. G., Sargent, F., Stanley, N. R., Berks, B. C., Robinson, C. and Palmer, T.** (1998). An Essential Component of a Novel Bacterial Protein Export System with Homologues in Plastids and Mitochondria. *Journal of Biological Chemistry*, **273**, 18003-18006.

**Bolhuis, A., Mathers, J. E., Thomas, J. D., Barrett, C. M. L. and Robinson, C.** (2001). TatB and TatC Form a Functional and Structural Unit of the Twin-arginine Translocase from Escherichia coli. *Journal of Biological Chemistry*, **276**, 20213-20219.

**Bottcher, B., Wynne, S. A. and Crowther, R. A.** (1997). Determination of the fold of the core protein of hepatitis B virus by electron cryomicroscopy. *Nature*, **386**, 88-91.

**Branston, S. D., Matos, C. F. R. O., Freedman, R. B., Robinson, C. and Keshavarz-Moore, E.** (2012). Investigation of the impact of Tat export pathway enhancement on E. coli culture, protein production and early stage recovery. *Biotechnology and Bioengineering*, **109**, 983-991.

**Brüser, T. and Sanders, C.** (2003). An alternative model of the twin arginine translocation system. *Microbiological Research*, **158**, 7-17.

**Buchanan, G., de Leeuw, E., Stanley, N. R., Wexler, M., Berks, B. C., Sargent, F. and Palmer, T.** (2002). Functional complexity of the twin-arginine translocase TatC component revealed by site-directed mutagenesis. *Mol Microbiol*, **43**, 1457-70.

**Chaddock, A. M., Mant, A., Karnauchov, I., Brink, S., Herrmann, R. G., Klösgen, R. B. and Robinson, C.** (1995). A new type of signal peptide: central role of a twin-arginine motif in transfer signals for the delta pH-dependent thylakoidal protein translocase. *EMBO journal*, **14**, 2715-2722.

**Chan, C. S., Zlomislic, M. R., Tieleman, D. P. and Turner, R. J.** (2007). The TatA subunit of Escherichia coli twin-arginine translocase has an N-in topology. *Biochemistry*, **46**, 7396-404.

**Christie, P. J., Atmakuri, K., Krishnamoorthy, V., Jakubowski, S. and Cascales, E.** (2005). Biogenesis, architecture, and function of bacterial type IV secretion systems. *Annu Rev Microbiol*, **59**, 451-85.

**Clark, S. A. and Theg, S. M.** (1997). A folded protein can be transported across the chloroplast envelope and thylakoid membranes. *Molecular Biology of the Cell*, **8**, 923-934.

**Cline, K., Ettinger, W. F. and Theg, S. M.** (1992). Protein-specific energy requirements for protein transport across or into thylakoid membranes. Two lumenal proteins are transported in the absence of ATP. *Journal of Biological Chemistry*, **267**, 2688-2696.

**Cline, K. and McCaffery, M.** (2007). Evidence for a dynamic and transient pathway through the TAT protein transport machinery. *EMBO J*, **26**, 3039 - 49.

**Cline, K. and Mori, H.** (2001). Thylakoid {Delta}pH-dependent precursor proteins bind to a cpTatC-Hcf106 complex before Tha4-dependent transport. *The Journal of Cell Biology*, **154**, 719-730.

**Cornelis, G. R.** (2006). The type III secretion injectisome. *Nat Rev Micro*, **4**, 811-825.

**Creighton, A. M., Hulford, A., Mant, A., Robinson, D. and Robinson, C.** (1995). A monomeric, tightly folded stromal intermediate on the delta pH-dependent thylakoidal protein transport pathway. *J Biol Chem*, **270**, 1663-9.

**Cristóbal, S., de Gier, J.-W., Nielsen, H. and von Heijne, G.** (1999). Competition between Sec- and TAT-dependent protein translocation in Escherichia coli. *EMBO journal*, **18**, 2982-2990.

**Dabney-Smith, C., Mori, H. and Cline, K.** (2006). Oligomers of Tha4 organize at the thylakoid Tat translocase during protein transport. *J Biol Chem*, **281**, 5476-83.

**De Buck, E., Vranckx, L., Meyen, E., Maes, L., Vandersmissen, L., Anne, J. and Lammertyn, E.** (2007). The twin-arginine translocation pathway is necessary for correct membrane insertion of the Rieske Fe/S protein in Legionella pneumophila. *FEBS Lett*, **581**, 259-64.

**De Leeuw, E., Granjon, T., Porcelli, I., Alami, M., Carr, S. B., Muller, M., Sargent, F., Palmer, T. and Berks, B. C.** (2002). Oligomeric properties and signal peptide binding by Escherichia coli Tat protein transport complexes. *J Mol Biol*, **322**, 1135-46.

**De Leeuw, E., Porcelli, I., Sargent, F., Palmer, T. and Berks, B. C.** (2001). Membrane interactions and self-association of the TatA and TatB components of the twin-arginine translocation pathway. *FEBS Lett*, **506**, 143-8.

**DeLisa, M. P., Samuelson, P., Palmer, T. and Georgiou, G.** (2002). Genetic Analysis of the Twin Arginine Translocator Secretion Pathway in Bacteria. *Journal of Biological Chemistry*, **277**, 29825-29831.

**DeLisa, M. P., Tullman, D. and Georgiou, G.** (2003). Folding quality control in the export of proteins by the bacterial twin-arginine translocation pathway. *Proceedings*

*of the National Academy of Sciences of the United States of America*, **100**, 6115-6120.

**Dilks, K., Rose, R. W., Hartmann, E. and Pohlschroder, M.** (2003). Prokaryotic utilization of the twin-arginine translocation pathway: a genomic survey. *Journal of Bacteriology*, **185**.

**Driessen, A. J.** (2001). SecB, a molecular chaperone with two faces. *Trends Microbiol*, **9**, 193-6.

**Dubini, A. and Sargent, F.** (2003). Assembly of Tat-dependent [NiFe] hydrogenases: identification of precursor-binding accessory proteins. In *FEBS Lett*, pp. 141-6. Edited by. Netherlands.

**Frank, J.** (2006). Three-Dimensional Electron Microscopy, second Edition: Oxford University Press).

**Frank, J., Shimkin, B. and Dowse, H.** (1981). Spider--A modular software system for electron image processing. *Ultramicroscopy*, **6**, 343-357.

**Frauenfeld, J., Gumbart, J., Sluis, E. O. v. d., Funes, S., Gartmann, M., Beatrix, B., Mielke, T., Berninghausen, O., Becker, T., Schulten, K. and Beckmann, R.** (2011). Cryo-EM structure of the ribosome–SecYE complex in the membrane environment. *Nat Struct Mol Biol*, **18**, 614-621.

**Fröbel, J., Rose, P. and Müller, M.** (2011). Early contacts between substrate proteins and TatA translocase component in twin-arginine translocation. *J Biol Chem*, **286**, 43679-89.

**Fröbel, J., Rose, P. and Müller, M.** (2012). Twin-arginine-dependent translocation of folded proteins. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **367**, 1029-1046.

**Genest, O., Seduk, F., Ilbert, M., Mejean, V. and Iobbi-Nivol, C.** (2006). Signal peptide protection by specific chaperone. *Biochem Biophys Res Commun*, **339**, 991-5.

**Gohlke, U., Pullan, L., McDevitt, C., Porcelli, I., de Leeuw, E., Palmer, T., Saibil, H. and Berks, B.** (2005). The TatA component of the twin-arginine protein transport system forms channel complexes of variable diameter. *Proc Natl Acad Sci U S A*, **102**, 10482-6.

**Gouffi, K., Gérard, F., Santini, C.-L. and Wu, L.-F.** (2004). Dual Topology of the Escherichia coli TatA Protein. *Journal of Biological Chemistry*, **279**, 11608-11615.

**Gouffi, K., Santini, C.-L. and Wu, L.-F.** (2002). Topology determination and functional analysis of the Escherichia coli TatC protein. *FEBS Letters*, **525**, 65-70.

**Haider, S., Hall, B. A. and Sansom, M. S. P.** (2006). Simulations of a protein translocation pore: SecY. *Biochemistry*, **45**.

**Harauz, G. and van Heel, M.** (1986). Exact filters for general geometry three dimensional reconstruction. *OPTIK.*, **73**, 146-156.

**Hartl, F. U., Lecker, S., Schiebel, E., Hendrick, J. P. and Wickner, W.** (1990). The binding cascade of SecB to SecA to SecY/E mediates preprotein targeting to the E. coli plasma membrane. *Cell*, **63**, 269-79.

**Hatzixanthis, K., Palmer, T. and Sargent, F.** (2003). A subset of bacterial inner membrane proteins integrated by the twin-arginine translocase. *Mol Microbiol*, **49**, 1377-90.

**Heikkila, M. P., Honisch, U., Wunsch, P. and Zumft, W. G.** (2001). Role of the Tat ransport system in nitrous oxide reductase translocation and cytochrome cd1 biosynthesis in Pseudomonas stutzeri. *J Bacteriol*, **183**, 1663-71.

**Henderson, I. R., Navarro-Garcia, F., Desvaux, M., Fernandez, R. C. and Ala'Aldeen, D.** (2004). Type V protein secretion pathway: the autotransporter story. *Microbiol Mol Biol Rev*, **68**, 692-744.

**Hicks, M. G., de Leeuw, E., Porcelli, I., Buchanan, G., Berks, B. C. and Palmer, T.** (2003). The Escherichia coli twin-arginine translocase: conserved residues of TatA and TatB family components involved in protein transport. *FEBS Letters*, **539**, 61-67.

**Hinsley, A. P., Stanley, N. R., Palmer, T. and Berks, B. C.** (2001). A naturally occurring bacterial Tat signal peptide lacking one of the [`]invariant' arginine residues of the consensus targeting motif. *FEBS Letters*, **497**, 45-49.

**Holzapfel, E., Eisner, G., Alami, M., Barrett, C. M. L., Buchanan, G., Luke, I., Betton, J. M., Robinson, C., Palmer, T., Moser, M. and Muller, M.** (2007). The Entire N-Terminal Half of TatC is Involved in Twin-Arginine Precursor Binding. *Biochemistry*, **46**, 2892-2898.

**Hu, Y., Zhao, E., Li, H., Xia, B. and Jin, C.** (2010). Solution NMR Structure of the TatA Component of the Twin-Arginine Protein Transport System from Gram-Positive Bacterium Bacillus subtilis. *Journal of the American Chemical Society*, **132**, 15942-15944.

**Hynds, P. J., Robinson, D. and Robinson, C.** (1998). The Sec-independent Twin-arginine Translocation System Can Transport Both Tightly Folded and Malfolded Proteins across the Thylakoid Membrane. *Journal of Biological Chemistry*, **273**, 34868-34874.

**Ignatova, Z., Hornle, C., Nurk, A. and Kasche, V.** (2002). Unusual signal peptide directs penicillin amidase from Escherichia coli to the Tat translocation machinery. *Biochem Biophys Res Commun*, **291**, 146-9.

**Ikeda, M. and Nakagawa, S.** (2003). The Corynebacterium glutamicum genome: features and impacts on biotechnological processes. *Appl Microbiol Biotechnol*, **62**, 99-109.

**Ilbert, M., Mejean, V., Giudici-Orticoni, M. T., Samama, J. P. and Iobbi-Nivol, C.** (2003). Involvement of a mate chaperone (TorD) in the maturation pathway of molybdoenzyme TorA. *J Biol Chem*, **278**, 28787-92.

**Ize, B., Gerard, F. and Wu, L. F.** (2002). In vivo assessment of the Tat signal peptide specificity in Escherichia coli. *Arch Microbiol*, **178**, 548-53.

**Ize, B., Gérard, F., Zhang, M., Chanal, A., Voulhoux, R., Palmer, T., Filloux, A. and Wu, L.-F.** (2002). In vivo dissection of the Tat translocation pathway in Escherichia coli. *Journal of Molecular Biology*, **317**, 327-335.

**Ize, B., Stanley, N. R., Buchanan, G. and Palmer, T.** (2003). Role of the Escherichia coli Tat pathway in outer membrane integrity. *Mol Microbiol*, **48**, 1183-93.

**Jack, R. L., Buchanan, G., Dubini, A., Hatzixanthis, K., Palmer, T. and Sargent, F.** (2004). Coordinating assembly and export of complex bacterial proteins. *EMBO journal*, **23**, 3962-3972.

**Jack, R. L., Sargent, F., Berks, B. C., Sawers, G. and Palmer, T.** (2001). Constitutive expression of Escherichia coli tat genes indicates an important role for the twin-arginine translocase during aerobic and anaerobic growth. *J Bacteriol*, **183**, 1801-4.

**Jakob, M., Kaiser, S., Gutensohn, M., Hanner, P. and Klosgen, R. B.** (2009). Tat subunit stoichiometry in Arabidopsis thaliana challenges the proposed function of TatA as the translocation pore. *Biochim Biophys Acta*, **1793**, 388-94.

**Jarvis, P. and Robinson, C.** (2004). Mechanisms of Protein Import and Routing in Chloroplasts. *Current Biology*, **14**, R1064-R1077.

**Johnson, T. L., Abendroth, J., Hol, W. G. and Sandkvist, M.** (2006). Type II secretion: from structure to function. *FEMS Microbiol Lett*, **255**, 175-86.

**Jongbloed, J. D. H., Grieger, U., Antelmann, H., Hecker, M., Nijland, R., Bron, S. and Van Dijl, J. M.** (2004). Two minimal Tat translocases in Bacillus. *Molecular Microbiology*, **54**, 1319-1325.

**Jongbloed, J. D. H., Martin, U., Antelmann, H., Hecker, M., Tjalsma, H., Venema, G., Bron, S., van Dijl, J. M. and Muller, J.** (2000). TatC Is a Specificity Determinant for Protein Secretion via the Twin-arginine Translocation Pathway. *Journal of Biological Chemistry*, **275**, 41350-41357.

**Jongbloed, J. D. H., van der Ploeg, R. and van Dijl, J. M.** (2006). Bifunctional TatA subunits in minimal Tat protein translocases. *Trends in Microbiology*, **14**, 2-4.

**Klösgen, R. B., Brock, I. W., Herrmann, R. G. and Robinson, C.** (1992). Proton gradient-driven import of the 16 kDa oxygen-evolving complex protein as the full precursor protein by isolated thylakoids. *Plant Molecular Biology*, **18**, 1031-1034.

**Knowles, T. J., Finka, R., Smith, C., Lin, Y. P., Dafforn, T. and Overduin, M.** (2009). Membrane proteins solubilized intact in lipid containing nanoparticles bounded by styrene maleic acid copolymer. *J Am Chem Soc*, **131**, 7484-5.

**Kostecki, J. S., Li, H., Turner, R. J. and DeLisa, M. P.** (2010). Visualizing interactions along the Escherichia coli twin-arginine translocation pathway using protein fragment complementation. *PLoS One*, **5**, e9225.

**Kumamoto, C. A. and Francetic, O.** (1993). Highly selective binding of nascent polypeptides by an Escherichia coli chaperone protein in vivo. *J Bacteriol*, **175**, 2184-8.

**Lange, C., Müller, S. D., Walther, T. H., Bürck, J. and Ulrich, A. S.** (2007). Structure analysis of the protein translocating channel TatA in membranes using a multi-construct approach. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, **1768**, 2627-2634.

**Leake, M. C., Greene, N. P., Godun, R. M., Granjon, T., Buchanan, G., Chen, S., Berry, R. M., Palmer, T. and Berks, B. C.** (2008). Variable stoichiometry of the TatA component of the twin-arginine protein transport system observed by in vivo single-molecule imaging. *Proceedings of the National Academy of Sciences*, **105**, 15376-15381.

**Lee, P. A., Buchanan, G., Stanley, N. R., Berks, B. C. and Palmer, T.** (2002). Truncation Analysis of TatA and TatB Defines the Minimal Functional Units Required for Protein Translocation. *The Journal of Bacteriology*, **184**, 5871-5879.

**Lee, P. A., Orriss, G. L., Buchanan, G., Greene, N. P., Bond, P. J., Punginelli, C., Jack, R. L., Sansom, M. S. P., Berks, B. C. and Palmer, T.** (2006). Cysteine-scanning Mutagenesis and Disulfide Mapping Studies of the Conserved Domain of the Twin-arginine Translocase TatB Component. *Journal of Biological Chemistry*, **281**, 34072-34085.

**Lee, P. A., Tullman-Ercek, D. and Georgiou, G.** (2006). The Bacterial Twin-Arginine Translocation Pathway. *Annual Review of Microbiology*, **60**, 373-395.

**Li, W., Schulman, S., Boyd, D., Erlandson, K., Beckwith, J. and Rapoport, T. A.** (2007). The plug domain of the SecY protein stabilizes the closed state of the translocation channel and maintains a membrane seal. *Mol Cell*, **26**, 511-21.

**Luirink, J. and Sinning, I.** (2004). SRP-mediated protein targeting: structure and function revisited. *Biochim Biophys Acta*, **1694**, 17-35.

**Luirink, J., von Heijne, G., Houben, E. and de Gier, J. W.** (2005). Biogenesis of inner membrane proteins in Escherichia coli. *Annu Rev Microbiol*, **59**, 329-55.

**Luke, I., Handford, J. I., Palmer, T. and Sargent, F.** (2009). Proteolytic processing of Escherichia coli twin-arginine signal peptides by LepB. *Arch Microbiol*, **191**, 919-25.

**Maldonado, B., Kneuper, H., Buchanan, G., Hatzixanthis, K., Sargent, F., Berks, B. C. and Palmer, T.** (2011). Characterisation of the membrane-extrinsic domain of the TatB component of the twin arginine protein translocase. *FEBS Lett*, **585**, 478-84.

**Mangels, D., Mathers, J., Bolhuis, A. and Robinson, C.** (2005). The Core TatABC Complex of the Twin-arginine Translocase in Escherichia coli: TatC Drives Assembly Whereas TatA is Essential for Stability. *Journal of Molecular Biology*, **345**, 415-423.

**Manting, E. H. and Driessen, A. J.** (2000). *Escherichia coli* translocase: the unravelling of a molecular machine. *Molecular Microbiology*, **37**, 226-238.

**Matos, C. F. R. O., Branston, S. D., Albiniak, A., Dhanoya, A., Freedman, R. B., Keshavarz-Moore, E. and Robinson, C.** (2012). High-yield export of a native heterologous protein to the periplasm by the tat translocation pathway in Escherichia coli. *Biotechnology and Bioengineering*, **109**, 2533-2542.

**McDevitt, C. A., Buchanan, G., Sargent, F., Palmer, T. and Berks, B. C.** (2006). Subunit composition and in vivo substrate-binding characteristics of Escherichia coli Tat protein complexes expressed at native levels. *FEBS J*, **273**, 5656-68.

**McDevitt, C. A., Hicks, M. G., Palmer, T. and Berks, B. C.** (2005). Characterisation of Tat protein transport complexes carrying inactivating mutations. *Biochem Biophys Res Commun*, **329**, 693-8.

**Mehner, D., Osadnik, H., Lunsdorf, H. and Bruser, T.** (2012). The Tat system for membrane translocation of folded proteins recruits the membrane-stabilizing Psp machinery in Escherichia coli. *The Journal of biological chemistry*, **287**, 27834-42.

**Mendel, S., McCarthy, A., Barnett, J. P., Eijlander, R. T., Nenninger, A., Kuipers, O. P. and Robinson, C.** (2008). The Escherichia coli TatABC System and a Bacillus subtilis TatAC-type System Recognise Three Distinct Targeting Determinants in Twin-arginine Signal Peptides. *Journal of Molecular Biology*, **375**, 661-672.

**Mindell, J. and Grigorieff, N.** (2003). Accurate determination of local defocus and specimen tilt in electron microscopy. *J Struct Biol*, **142**, 334-47.

**Molik, S., Karnauchov, I., Weidlich, C., Herrmann, R. G. and Klosgen, R. B.** (2001). The Rieske Fe/S protein of the cytochrome b6/f complex in chloroplasts: missing link in the evolution of protein transport pathways in chloroplasts? *J Biol Chem*, **276**, 42761-6.

**Mori, H. and Cline, K.** (2002). A twin arginine signal peptide and the pH gradient trigger reversible assembly of the thylakoid {Delta}pH/Tat translocase. *The Journal of Cell Biology*, **157**, 205-210.

**Mori, H. and Ito, K.** (2001). The Sec protein-translocation pathway. *Trends in Microbiology*, **9**, 494-500.

**Mould, R. M. and Robinson, C.** (1991). A proton gradient is required for the transport of two lumenal oxygen- evolving proteins across the thylakoid membrane. *Journal of Biological Chemistry*, **266**, 12189-12193.

**Mould, R. M., Shackleton, J. B. and Robinson, C.** (1991). Transport of proteins into chloroplasts. Requirements for the efficient import of two lumenal oxygen-evolving complex proteins into isolated thylakoids. *Journal of Biological Chemistry*, **266**, 17286-17289.

**Muñiz, M., Morsomme, P. and Riezman, H.** (2001). Protein Sorting upon Exit from the Endoplasmic Reticulum. *Cell*, **104**, 313-320.

**Müller, M. and Klösgen, R. B.** (2005). The Tat pathway in bacteria and chloroplasts (Review). *Mol Membr Biol*, **22**, 113-121.

**Müller, S. D., De Angelis, A. A., Walther, T. H., Grage, S. L., Lange, C., Opella, S. J. and Ulrich, A. S.** (2007). Structural characterization of the pore forming protein TatAd of the twin-arginine translocase in membranes by solid-state 15N-NMR. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, **1768**, 3071-3079.

**Nagai, K., Oubridge, C., Kuglstatter, A., Menichelli, E., Isel, C. and Jovine, L.** (2003). Structure, function and evolution of the signal recognition particle. *EMBO J*, **22**, 3479-85.

**Natale, P., Brüser, T. and Driessen, A. J. M.** (2008). Sec- and Tat-mediated protein secretion across the bacterial cytoplasmic membrane--Distinct translocases and mechanisms. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, **1778**, 1735-1756.

**Nolandt, O. V., Walther, T. H., Roth, S., Bürck, J. and Ulrich, A. S.** (2009). Structure analysis of the membrane protein TatCd from the Tat system of B. subtilis by circular dichroism. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, **1788**, 2238-2244.

**Oates, J., Barrett, C. M. L., Barnett, J. P., Byrne, K. G., Bolhuis, A. and Robinson, C.** (2005). The Escherichia coli Twin-arginine Translocation Apparatus Incorporates a Distinct Form of TatABC Complex, Spectrum of Modular TatA Complexes and Minor TatAB Complex. *Journal of Molecular Biology*, **346**, 295-305.

**Oates, J., Mathers, J., Mangels, D., Kühlbrandt, W., Robinson, C. and Model, K.** (2003). Consensus Structural Features of Purified Bacterial TatABC Complexes. *Journal of Molecular Biology*, **330**, 277-286.

**Oresnik, I. J., Ladner, C. L. and Turner, R. J.** (2001). Identification of a twin-arginine leader-binding protein. *Mol Microbiol*, **40**, 323-31.

**Orriss, G. L., Tarry, M. J., Ize, B., Sargent, F., Lea, S. M., Palmer, T. and Berks, B. C.** (2007). TatBC, TatB, and TatC form structurally autonomous units within the twin arginine protein transport system of Escherichia coli. *FEBS Letters*, **581**, 4091-4097.

**Panahandeh, S., Maurer, C., Moser, M., DeLisa, M. P. and Muller, M.** (2008). Following the path of a twin-arginine precursor along the TatABC translocase of Escherichia coli. *J. Biol. Chem.*, **283**, 33267-33275.

**Park, E. and Rapoport, T. A.** (2011). Preserving the membrane barrier for small molecules during bacterial protein translocation. *Nature*, **473**, 239-42.

**Penczek, P. A., Grassucci, R. A. and Frank, J.** (1994). The ribosome at improved resolution: new techniques for merging and orientation refinement in 3D cryo-electron microscopy of biological particles. *Ultramicroscopy*, **53**, 251.

**Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. and Ferrin, T. E.** (2004). UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem*, **25**, 1605-12.

**Pop, O., Martin, U., Abel, C. and Muller, J. P.** (2002). The Twin-arginine Signal Peptide of PhoD and the TatAd/Cd Proteins of Bacillus subtilis Form an Autonomous Tat Translocation System. *Journal of Biological Chemistry*, **277**, 3268-3273.

**Pop, O. I., Westermann, M., Volkmer-Engert, R., Schulz, D., Lemke, C., Schreiber, S., Gerlach, R., Wetzker, R. and Muller, J. P.** (2003). Sequence-specific Binding of prePhoD to Soluble TatAd Indicates Protein-mediated Targeting of the Tat Export in Bacillus subtilis. *Journal of Biological Chemistry*, **278**, 38428-38436.

**Porcelli, I., de Leeuw, E., Wallis, R., van den Brink-van der Laan, E., de Kruijff, B., Wallace, B. A., Palmer, T. and Berks, B. C.** (2002). Characterization and Membrane Assembly of the TatA Component of the Escherichia coli Twin-Arginine Protein Transport System. *Biochemistry*, **41**, 13690-13697.

**Pugsley, A. P.** (1993). The complete general secretory pathway in gram-negative bacteria. *Microbiol Rev*, **57**, 50-108.

**Punginelli, C., Maldonado, B., Grahl, S., Jack, R., Alami, M., Schroder, J., Berks, B. C. and Palmer, T.** (2007). Cysteine Scanning Mutagenesis and Topological Mapping of the Escherichia coli Twin-Arginine Translocase TatC Component. *The Journal of Bacteriology*, **189**, 5482-5494.

**Radermacher, M.** (1988). Three-Dimensional reconstruction of single particles from random and nonrandom tilt series. *Journal of Electron Microscopy Technique*, **9**, 359-394.

**Radermacher, M., Wagenknecht, T., Verschoor, A. and Frank, J.** (1987). Three-dimensional reconstruction from a single-exposure, random conical tilt series applied to the 50S ribosomal subunit of Escherichia coli. *J Microsc*, **146**, 113-36.

**Radermacher, M., Wagenknecht, T., Verschoor, A. and Frank, J.** (1987). Three-dimensional structure of the large ribosomal subunit from Escherichia coli. *EMBO J*, **6**, 1107-14.

**Randall, L. L. and Hardy, S. J.** (1986). Correlation of competence for export with lack of tertiary structure of the mature species: a study in vivo of maltose-binding protein in E. coli. *Cell*, **46**, 921-8.

**Robinson, C. and Bolhuis, A.** (2004). Tat-dependent protein targeting in prokaryotes and chloroplasts. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, **1694**, 135-147.

**Robinson, C., Matos, C. F. R. O., Beck, D., Ren, C., Lawrence, J., Vasisht, N. and Mendel, S.** (2011). Transport and proofreading of proteins by the twin-arginine translocation (Tat) system in bacteria. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, **1808**, 876-884.

**Rodrigue, A., Chanal, A., Beck, K., Muller, M. and Wu, L. F.** (1999). Co-translocation of a periplasmic enzyme complex by a hitchhiker mechanism through the bacterial tat pathway. *J Biol Chem*, **274**, 13223-8.

**Rubinstein, J. L.** (2007). Structural analysis of membrane protein complexes by single particle electron microscopy. *Methods*, **41**, 409-416.

**Sambasivarao, D., Turner, R. J., Simala-Grant, J. L., Shaw, G., Hu, J. and Weiner, J. H.** (2000). Multiple roles for the twin arginine leader sequence of dimethyl sulfoxide reductase of Escherichia coli. *J Biol Chem*, **275**, 22526-31.

**Sanders, C., Wethkamp, N. and Lill, H.** (2001). Transport of cytochrome c derivatives by the bacterial Tat protein translocation system. In *Mol Microbiol*, pp. 241-6. Edited by. England.

**Santini, C.-L., Ize, B., Chanal, A., Müller, M., Giordano, G. and Wu, L.-F.** (1998). A novel Sec-independent periplasmic protein translocation pathway in Escherichia coli. *EMBO journal*, **17**, 101-112.

**Sargent, F., Berks, B. and Palmer, T.** (2002). Assembly of membrane-bound respiratory complexes by the Tat protein-transport system. *Archives of Microbiology*, **178**, 77-84.

**Sargent, F., Bogsch, E. G., Stanley, N. R., Wexler, M., Robinson, C., Berks, B. C. and Palmer, T.** (1998). Overlapping functions of components of a bacterial Sec-independent protein export pathway. *EMBO journal*, **17**, 3640-3650.

**Sargent, F., Gohlke, U., De Leeuw, E., Stanley, N., Palmer, T., Saibil, H. and Berks, B.** (2001). Purified components of the Escherichia coli Tat protein transport system form a double-layered ring structure. *Eur J Biochem*, **268**, 3361-7.

**Sargent, F., Stanley, N. R., Berks, B. C. and Palmer, T.** (1999). Sec-independent Protein Translocation in Escherichia coli. A DISTINCT AND PIVOTAL ROLE FOR THE TatB PROTEIN. *Journal of Biological Chemistry*, **274**, 36073-36082.

**Schaerlaekens, K., Schierova, M., Lammertyn, E., Geukens, N., Anne, J. and Van Mellaert, L.** (2001). Twin-Arginine Translocation Pathway in Streptomyces lividans. *The Journal of Bacteriology*, **183**, 6727-6732.

**Schreiber, S., Stengel, R., Westermann, M., Volkmer-Engert, R., Pop, O. I. and Müller, J. P.** (2006). Affinity of TatCd for TatAd Elucidates Its Receptor Function in the Bacillus subtilis Twin Arginine Translocation (Tat) Translocase System. *Journal of Biological Chemistry*, **281**, 19977-19984.

**Settles, A. M., Yonetani, A., Baron, A., Bush, D. R., Cline, K. and Martienssen, R.** (1997). Sec-Independent Protein Translocation by the Maize Hcf106 Protein. *Science*, **278**, 1467-1470.

**Shaikh, T. R., Gao, H., Baxter, W. T., Asturias, F. J., Boisset, N., Leith, A. and Frank, J.** (2008). SPIDER image processing for single-particle reconstruction of biological macromolecules from electron micrographs. *Nat. Protocols*, **3**, 1941-1974.

**Stanley, N. R., Palmer, T. and Berks, B. C.** (2000). The Twin Arginine Consensus Motif of Tat Signal Peptides Is Involved in Sec-independent Protein Targeting in Escherichia coli. *Journal of Biological Chemistry*, **275**, 11591-11596.

**Tarry, M. J., Schäfer, E., Chen, S., Buchanan, G., Greene, N. P., Lea, S. M., Palmer, T., Saibil, H. R. and Berks, B. C.** (2009). Structural analysis of substrate binding by the TatBC component of the twin-arginine protein transport system. *Proceedings of the National Academy of Sciences*, **106**, 13284-13289.

**Tian, P. and Andricioaei, I.** (2006). Size, motion, and function of the SecY translocon revealed by molecular dynamics simulations with virtual probes. *Biophysical Journal*, **90**.

**Tjalsma, H., Bolhuis, A., Jongbloed, J. D. H., Bron, S. and van Dijl, J. M.** (2000). Signal Peptide-Dependent Protein Transport in Bacillus subtilis: a Genome-Based Survey of the Secretome. *Microbiology and Molecular Biology Reviews*, **64**, 515-547.

**Valent, Q. A.** (2001). Signal recognition particle mediated protein targeting in Escherichia coli. *Antonie Van Leeuwenhoek*, **79**, 17-31.

**van der Ploeg, R., Barnett, J. P., Vasisht, N., Goosens, V. J., Pother, D. C., Robinson, C. and van Dijl, J. M.** (2011). Salt sensitivity of minimal twin arginine translocases. *J Biol Chem*, **286**, 43759-70.

**Walker, M. B., Roy, L. M., Coleman, E., Voelker, R. and Barkan, A.** (1999). The Maize tha4 Gene Functions in Sec-Independent Protein Transport in Chloroplasts and Is Related to hcf106, tatA, and tatB. *The Journal of Cell Biology*, **147**, 267-276.

**Walther, T. H., Grage, S. L., Roth, N. and Ulrich, A. S.** (2010). Membrane Alignment of the Pore-Forming Component TatAd of the Twin-Arginine Translocase from Bacillus subtilis Resolved by Solid-State NMR Spectroscopy. *Journal of the American Chemical Society*, **132**, 15945-15956.

**Ward Jr, J. H.** (1963). Hierarchical grouping to optimize an objective function. *Journal of the American statistical association*, 236-244.

**Weiner, J. H., Bilous, P. T., Shaw, G. M., Lubitz, S. P., Frost, L., Thomas, G. H., Cole, J. A. and Turner, R. J.** (1998). A Novel and Ubiquitous System for Membrane Targeting and Secretion of Cofactor-Containing Proteins. *Cell*, **93**, 93-101.

**Westermann, M., Pop, O. I., Gerlach, R., Appel, T. R., Schlörmann, W., Schreiber, S. and Müller, J. P.** (2006). The TatAd component of the Bacillus subtilis twin-arginine protein transport system forms homo-multimeric complexes in its cytosolic and membrane embedded localisation. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, **1758**, 443-451.

**Wexler, M., Sargent, F., Jack, R. L., Stanley, N. R., Bogsch, E. G., Robinson, C., Berks, B. C. and Palmer, T.** (2000). TatD Is a Cytoplasmic Protein with DNase Activity. NO REQUIREMENT FOR TatD FAMILY PROTEINS IN Sec-INDEPENDENT PROTEIN EXPORT. *Journal of Biological Chemistry*, **275**, 16717-16722.

**White, H. E., Saibil, H. R., Ignatiou, A. and Orlova, E. V.** (2004). Recognition and Separation of Single Particles with Size Variation by Statistical Analysis of their Images. *Journal of Molecular Biology*, **336**, 453-460.

**Widdick, D. A., Eijlander, R. T., van Dijl, J. M., Kuipers, O. P. and Palmer, T.** (2008). A facile reporter system for the experimental identification of twin-arginine translocation (Tat) signal peptides from all kingdoms of life. *J Mol Biol*, **375**, 595-603.

**Yahr, T. L. and Wickner, W. T.** (2001). Functional reconstitution of bacterial Tat translocation in vitro. *EMBO journal*, **20**, 2472-2479.

**Yen, M.-R., Tseng, Y.-H., Nguyen, E., Wu, L.-F. and Saier, M.** (2002). Sequence and phylogenetic analyses of the twin-arginine targeting (Tat) protein export system. *Archives of Microbiology*, **177**, 441-450.

# *Chapter 8*
# *Published work*

# *Appendix:*

# *SPIDER batch files*

**T3D_tif_to_spi.cor - Convert TIF format files to SPIDER files**

```
; ----------- Parameters --------------

x10 = 4096 ; Image dimensions (in pixels)
x13 = 16384 ; Header Bytes in TIF image

; ----------- Input files --------------

FR L
[FILENUMS]  ; file numbers

FR L
[mic_tif]Micro_{**X12}   ; micrographs to convert

; ----------- Output files --------------
FR L
[mic_spi]Micro_{**X12}   ; spider output micrographs

; -------------- END BATCH HEADER -----------------

; get the numbers of files
UD N,x20
[FILENUMS]

DO LB1 x11=1,X20 ; loop through micrographs
UD x11,x12
[FILENUMS]

CP FROM RAW ; copy to SPIDER format
[mic_tif]
(8)
x10,x10
x13
[mic_spi]

LB1

EN D
```

**T3D_power.spi - Generate Power Spectra**

```
; ----------- Parameters --------------

x71 = 500  ; window size of small pieces (Sx=Sy)
x72 = 20   ; % of the overlap in x
x73 = 20   ; % of the overlap in y
x74 = 20   ; dist. from the edge (x)
x75 = 20   ; dist. from the edge (y)
x89 = 1    ; 1 = mask out centre of spectrum; 0 = don't mask
x88 = 1    ; decimation factor

; ----------- Input files --------------
FR L
[params]      ; parameter file
```

```
FR L
[FILENUMS]     ; file numbers

FR L
[mic]Micro_{***x12}   ; micrographs

; ----------- Output files --------------

FR L
[outdir]   ; output directory

FR L
[spectrum][outdir]/pw_avg{****x12}   ; power spectra

FR L
[roo][outdir]/roo{****x12}   ; rotational average (doc file)

; -------------- END BATCH HEADER --------------------------
;
; This gets the file numbers then calls power_p1 in the main loop

; get the number of files
UD N,x20
[FILENUMS]

VM                      ; Make sure output dir. present
mkdir -p [outdir]

; create an in-core doc file array
FR G
[tmparr]tmparray
SD IC NEW
[tmparr]
(4,x71/2)   ; creates a 4 column doc file

DO LB1 x11=1,x20
UD x11,x12
[FILENUMS]
; x12 is now the file number

VM
echo Working on micrograph {**x12}

    CP    ; copy micrograph to _1 variable for power-p1
    [mic]
    _1


; if the output doc files already exist, delete them
DE
[roo]

@power-p1[x71,x72,x73,x74,x75,x89]
_1
[spectrum] ; output image
[roo]      ; output doc file
[tmparr]


LB1
```

```
EN D
```

## T3D_defocus.spi - Calculate untilted micrograph defocus

```
; ----------- Input files --------------

FR G
[pow]pw_avg{****x31}   ; 2D power spectra

FR G
[FILENUMS]    ; file numbers

FR L
[params]    ; parameter file

; ----------- Output files --------------

FR G
[ctf]ctf{***x31}   ; output doc files with spectrum, envelope, noise

FR G
[out]defocusdoc    ; doc file of defocus values for all inputs

; -------------- END BATCH HEADER -------------------------

DE
[out]

SD /    micrograph   defocus     astig.ang   astig.mag  cutoff.freq
[out]

; make sure all parameters are calculated
@calcparms
[params]

UD 5,x55   ; pixel size
[params]
UD 7,x56   ; spherical aberration
[params]
UD 12,x57  ; amplitude contrast ratio
[params]
UD 14,x58  ; lambda
[params]

UD N,x30
[FILENUMS]

DO LB1 x11 = 1, x30

   UD x11,x31
   [FILENUMS]    ; x31 is file number

   DE
   [ctf]

   TF ED,X12,X13,X14,X15,X16
   [pow]      ; input 2D spectrum
```

```
   (x55, x56) ; pixel size (A), spherical aberration (mm)
   (x58)      ; lambda
   (x57)      ; ampl. contrast ratio
   [ctf]      ; output doc file

   SD x11,x31,x14,x12,x13,x16
   [out]

LB1

EN D
```

**T3D_TiltDefocus.fed – Calculate tilted particle defocus**

```
;-------------------------------------------------------------!
; tiltdefocus.fed/hbl : batch for calculating individual particle
defocus based on CTFTILT output :
;-------------------------------------------------------------!

; ---------- Parameters -----------
x18 = 1            ; Micrograph number
x61 = 2048         ; Central x coordinate
x62 = 2048         ; Central y coordinate
x63 = 0.952117     ; N1 value
x64 = -0.305733    ; N2 value
x65 = 2.62         ; Pixel size (A)
x66 = 46.42        ; Tilt Angle (TANGLE)
x72 = 13350.77     ; DFMID1
x73 = 12231.80     ; DFMID2
x74 = 11.68        ; angle of astigmatism


; ------------- Input files -----------

FR L
[ptcl_list]nparticles_{***x18}   ; particle-by-micrograph lists

FR L
[init_tilted_coords] /dct{***x18} ; tilted ptcl coords(from JWEB)

; ------------- Output files -------------

FR L
[tilted_defocus]defocustilt{***x18} ; tilted ptcl defocus values

; ---------------- END BATCH HEADER ------------------

MD
TR ON

UD N x12
[ptcl_list]

DO LB1 x13=1,x12

UD x13,x10
```

```
[ptcl_list]

UD x13,x10,x41,x42
[init_tilted_coords]


x51=x61-x41 ; DX=CX-NX
x52=x62-x42 ; DY=CY-NY ****MODIFIED****

;  DF = (N1*DX+N2*DY)*PSIZE*TAN(TANGLE)

x71 = (x63*x51+x64*x52)*x65*TAN(x66)

;  DFL1  = DFMID1 +DF
x81=x72+x71
;  DFL2  = DFMID2 +DF
x82=x73+x71

SD x13,x10,x81,x82,x74
[tilted_defocus]

UD E

LB1

vm
echo "Done"; date

EN
```

**T3D_mv_microdocs.spi - Move jweb document files to a new directory**

```
; MOVE TILTED MICRO COORDINATE DOCS TO DOC DIR.
;
; ---------------- Parameters ----------------

x10 = 10 ; number of tilt pairs

; --------------- Input files ---------------

FR L
[micro_dir]spi ; spider format micrograph dir.

FR L
[doc_dir]doc ; output document dir.

; ------------- END BATCH HEADER -------------

vm
echo "if(! -d [doc_dir]) mkdir -p [doc_dir]"|csh

DO LB1 x12=1,x10

vm
cp [micro_dir]/P_{**x12}/dcb001.spi [doc_dir]/dcb001.spi
```

```
vm
cp [micro_dir]/P_{**x12}/dct001.spi [doc_dir]/dct001.spi
vm
cp [micro_dir]/P_{**x12}/dcu001.spi [doc_dir]/dcu001.spi

LB1

vm
echo "Done"; date

EN D
```

**T3D_Tilt_box-out.fed - Window out particles from tilted micrographs**

```
; ----- Parameters -----
x76 = 22    ; last micrograph number
x81 = 128.  ; image dimension
x82=x81/2.  ; 1/2 image dimension

; ---- Input files ----
FR L
[init_tilt_coords]P_{***x18}/dct{***x18} ; tilted image coords

FR L
[tilt_micrograph]P_{**x18}/T_Micro_{**x18}  ; tilted micrograph

; ---- Output files ----
FR L
[ptcl_dir]Particles/Tilt ; particle directory

FR L
[tilt_images][ptcl_dir]/tilt{*****x11}; windowed, tilted images

FR L
[doc_dir]doc   ; document directory

FR L
[new_tilt_coords][doc_dir]/dwintilt{***x18} ; new tilt img coords

; ---------------- END BATCH HEADER -----------------------------

MD
TR OFF

VM
echo "if(! -d [ptcl_dir]) mkdir -p [ptcl_dir]"|csh

VM
echo "if(! -d [doc_dir]) mkdir -p [doc_dir]"|csh

; Starts the do loop of the boxing

DO LB1 x18=1,x76

DOC REN ; FIRST RE-KEYS THE DOC FILES SO EACH STARTS WITH 1 BUT
        ; KEEPS THE SEQUENTIAL PTCL NO. IN 1ST COLUMN
```

```
      UD N x17
      [init_tilt_coords]

            DO LB2 x10=1,x17

;   reads central X and Y coords for each ptcl in the micrograph
            UD x10,x11,x12,x13,x14,x15,x16
            [init_tilt_coords]

;   subtract 1/2 image dimension from centre coordinates
            X12=X12-x82
            X13=X13-x82

;   boxing out of each particle in a set of small images
            WI
            [tilt_micrograph]
            [tilt_images]
            X81,X81
            X12,X13

;   save windowing (top left) coordinates in a new document file
            SD x10,x11,x12,x13
            [new_tilt_coords]

            LB2

      LB1

EN
```

**T3D_Untilt_box-out.fed - Window out particles from untilted micrograph**

```
; ----- Parameters -----
x76 = 22    ; last micrograph number
x81 = 128   ; image dimension
x82=x81/2.  ; 1/2 image dimension

; ---- Input files ----

FR l
[init_untilt_coords]P_{**x18}/dcu{***x18} ; untilted image coords

FR L
[untilt_micrograph]P_{**x18}/Micro_{**x18} ; untilted micrograph

; ---- Output files ----

FR G
[u_img_dir]Particles/Untilt   ; directory for windowed particles

FR L
[untilt_images][ptcl_dir]/unt{*****x11}; windowed,untilted images

FR G
[coord_dir]doc   ; directory for IMC
```

```
FR L
[new_untilt_coords][doc_dir]/dwinunt{***x18}  ; new untilt image
                                              ; coords

FR G
[ptcl_list_dir]ptcl_lists   ; directory for particle lists

FR L
[ptcl_by_micro_list][ptcl_list_dir] ; template for non-sequential
                                    ; particle lists
FR L
[ptcl_series_list][ptcl_list_dir] ; list of total ptcls in series

; -- END BATCH HEADER --

MD
TR OFF

; Starts the do loop of the boxing through the micrographs :

VM
echo "if(! -d [u_img_dir]) mkdir -p [u_img_dir]"|csh

VM
echo "if(! -d [coord_dir]) mkdir -p [coord_dir]"|csh

VM
echo "if(! -d [ptcl_list_dir]) mkdir -p [ptcl_list_dir]"|csh


DO LB1 x18=1,x76

      DOC REN ; FIRST RE-KEYS THE DOC FILES SO EACH STARTS WITH 1
              ; BUT KEEPS THE SEQUENTIAL PTCL NO. IN 1ST COLUMN
      [init_untilted_coords]
        [init_untilted_coords]

      UD N x17
      [init_untilted_coords]

            DO LB2 x10=1,x17

;reads central X and Y coords for each ptcl in the micrograph
              UD x10,x11,x12,x13,x14,x15,x16
            [init_untilted_coords]

;subtract 1/2 image dimension from center coordinates
            X12=X12-x82
            X13=X13-x82

;boxing out of each particle in a continuous set of small images
            WI
            [untilted_micrograph]
            [untilted_images]
            x81,x81
            X12,X13
```

```
;save windowing (top left) coordinates in a new document file
            SD x10,x11,x12,x13
            [new_untilted_coords]

            SD x10,x11
            [ptcl_by_micro_list]_{***x18}
            LB2
      LB1

DOC COMBINE
[ptcl_by_micro_list]_***
1-x76
[ptcl_series_list]

EN
```

**T3D_normalise.prj - Normalise particles**

```
; ---- Input files ----
FR G
[particle_list_dir]ptcl_lists ; directory for particle list

FR G
[particle_list]listparticles   ; name for particle list

FR L
[ptcl_series] ; template for series of particles

; ---- Output files ----
FR L
[norm_dir]norm ; directory for normalised particles

FR L
[norm_particles][norm_dir]/N_tilt   ; template for norm ptcls

; -------------- END BATCH HEADER -------------

MD
TR OFF

VM
echo "if(! -d [norm_dir]) mkdir -p [norm_dir]"|csh

UD N,X30,X31
[particle_list]

; loop through particles
DO LB1 x32=1,X30
    ; Normalize windowed particles
    FS x40,x41,x42,x43             ; Find statistics
    [ptcl_series]{*****X32}
    AR                             ; Normalize
    [ptcl_series]{*****X32}
    [norm_particles]{*****X32}
    (P1-x42)/x43
```

```
LB1

VM
echo "Done"; date

EN
```

**T3D_ctfcorr.cor - CTF-correct particles**

```
; -------------- Parameters --------------
x56 = 1  ; progress-message displayed every Nth micrograph

; ------------- Input files -------------
FR L
[parameter_doc]params   ; parameter doc
FR G
[defocus_by_micrograph]defocusdoc   ; defocus doc
FR L
[particles_by_micrograph]  ; particle-by-micrograph list
FR L
[particles]   ; particles

; ------------- Output files -------------
FR G
[ctf_dir]ctf/Untilt   ; output directory
FR G
[ctf_corrected][ctf_dir]/NC_unt*****   ; CTF-corrected particles

; ----------- END BATCH HEADER -----------

; get speherical aberration
UD 7,x16
[parameter_doc]

; get amplitude contrast
UD 12,x15
[parameter_doc]

; get wavelength
UD 14,x25
[parameter_doc]

; get maximum spatial frequency
UD 15,x18
[parameter_doc]

; get window size
UD 17,x17
[parameter_doc]

UD E  ; close doc


VM
echo "CTF-correcting particles"; date
```

```
VM
echo "if(! -d [ctf_dir]) mkdir [ctf_dir]"|csh

; get #micrographs
UD N x29
[defocus_by_micrograph]

; loop through micrographs
DO LB1 x11=1,x29
; get micrograph# (x10), defocus value (Angstroms,x14),
; astig and astig angle

    UD IC,x11,x10,x14,x20,x21
    [defocus_by_micrograph]

    IF (int(x11/x56).eq.x11/x56) THEN
        VM
echo "Working on micrograph #{***x10} ({***x11} out of {***x29})"
    ENDIF

    ; calculate (binary) CTF
    TF CT
    _1
    x16         ; spherical aberration
    x14,x25     ; defocus, wavelength
    x17,x17     ; x,y-dimensions
    x18         ; maximum spatial frequency
    (0.005,0)   ; source size, defocus spread
    (x21,x20)   ; astigmatism, azimuth
    x15         ; amplitude contrast
    (-1)        ; sign

    ; get #particles
    UD N x12
    [particles_by_micrograph]x10
    ; x10 = micrograph#

    ; loop through particles
    DO LB2 x13=1,x12
        ; get particle# (x19)
        UD IC x13,x19
        [particles_by_micrograph]x10
        ; x10 = micrograph#

        ; fourier transform each particle
        FT
        [particles]x19
        _3

        ; divide each particle's FT by the appropriate CTF
        MU
        _3  ; image FT
        _1  ; CTF
        _4  ; OUTPUT
        *   ; no further images to multiply

        ; inverse fourier transform each CTF-corrected FT
```

```
        FT
        _4
        [ctf_corrected]x19
    LB2
    ; end particle-loop

    UD ICE  ; close doc
    [particles_by_micrograph]x10
LB1
; end micrograph-loop

UD ICE  ; close doc
[defocus_by_micrograph]

EN D
```

**T3D_tiltctfcorr.cor - CTF-correct tilted particles**

```
; -------------- Parameters --------------

x56 = 1  ; progress-message displayed every Nth micrograph
x54 = 10 ; number of micrographs

; ------------- Input files -------------
FR L
[parameter_doc]../params   ; parameter doc

FR L
[tilted_defocus]doc/defocustilt*** ; defocus by particle doc

FR L
[particles]Particles/Tilt/N_tilt*****   ; particles

; ------------- Output files -------------
FR L
[ctf_dir]ctf/Tilt   ; output directory

FR L
[ctf_corrected][ctf_dir]/NC_tilt*****   ; CTF-corrected particles

; ----------- END BATCH HEADER -----------

; get spherical aberration
UD 7,x14
[parameter_doc]

; get amplitude contrast
UD 12,x15
[parameter_doc]

; get wavelength
UD 14,x25
[parameter_doc]

; get maximum spatial frequency
```

```
UD 15,x18
[parameter_doc]

; get window size
UD 17,x17
[parameter_doc]

UD E  ; close doc


VM
echo "CTF-correcting particles"; date

VM
echo "if(! -d [ctf_dir]) mkdir [ctf_dir]"|csh


; loop through micrographs
DO LB1 x28=1,x54

    ; get number of particles (x10)
    UD N x10
    [tilted_defocus]x28

  DO LB2 x13 = (1),x10

    ; get particle number (x11), defocus values, and astig angle
    UD x13,x11,x31,x32,x37
    [tilted_defocus]x28

    ; convert defocus values into absolute values
    x33 = abs(x31)
    x34 = abs(x32)

    ; use the avg of DFL1 and DFL2 as final defocus value (x35)
    x35 = ((x33)+(x34))/(2)

    ; The difference between DF1 and DF2 gives the variation in
    ; defocus due to axial astigmatism
    x36 = (x31)-(x32)

    ; correct for negative astig values and covert angle astig
    ; into spider convention from mrc.

    x38 = x36

    IF (x38.LE.0) THEN
    x36 = -(x38)
    x37 = x37+(45)

    ELSE
    x37 = x37-(45)

    ENDIF

        VM
        echo "Working on micrograph #{***x28} out of {***x54}
              particle #{***x13} out of {***x10}"
```

```
    ; calculate (binary) CTF
    TF CT
    _1
    x14         ; spherical aberration
    x35,x25     ; defocus, wavelength
    x17,x17     ; x,y-dimensions
    x18         ; maximum spatial frequency
    (0.005,0)   ; source size, defocus spread
    (x36,x37)   ; astigmatism, azimuth
    x15         ; amplitude contrast
    (-1)        ; sign

        ; fourier transform each particle
        FT
        [particles]x11
        _3

        ; divide each particle's FT by the appropriate CTF
        MU
        _3  ; image FT
        _1  ; CTF
        _4  ; OUTPUT
        *   ; no further images to multiply

        ; inverse fourier transform each CTF-corrected FT
        FT
        _4
        [ctf_corrected]x11
    LB2
    ; end particle-loop

    ;UD E  ; close doc
LB1
; end micrograph-loop

EN
```

**T3D_filter.cor - Filter particles**

```
; LOW-PASS FILTER IMAGES

; ------------- Parameters -------------
x10 =    ; Butterworth pass-band
x14 =    ; Butterworth stop-band
x56 =    ; progress-message displayed every Nth particle

; ------------- Input files -------------
FR G
[group_particles] ; group-particle list
FR G
[ctf_dir]ctf   ; CTF directory
FR G
[ctf_corrected][ctf_dir]/NC_ptcl*****   ; CTF-corrected particles
```

```
; ------------ Output files ------------
FR G
[filtered_dir]flt   ; output directory
FR G
[filtered][filtered_dir]/NCF_ptcl*****   ; filtered particles

; ---------- END BATCH HEADER ----------

VM
echo "if(! -d [filtered_dir]) mkdir -r [filtered_dir]"|csh

VM
echo "Filtering particles"; date

; get #particles
UD N x11
[group_particles]

DO LB1 x12=1,x11
    ; get particle#
    UD IC x12,x19
    [group_particles]

    IF (int(x12/x56).eq.x12/x56) THEN
        VM
        echo "Working on particle #{*****x19} ({*****x19} out of
              {*****x11})"
    ENDIF

    FQ
    [ctf_corrected]x19
    [filtered]x19
    (7)      ; Butterworth filter
    x10,x14  ; passband, stopband
LB1

UD ICE  ; close doc
[group_particles]

VM
echo "Done"; date

EN D
```

**T3D_centre.fed - Centre particles**

```
MD
TR OFF

; ----- Parameters -----
```

```
x50 = 128        ; image dimension
x51 = x50/2+1    ; image centre coordinates
x60 = 20         ; mask radius
x80 = 0          ; mask final particles? (1 = yes)
x90 = 20         ; final mask radius
x92 = 1          ; subtract CG? (1 = yes)
x95 = 0          ; use custom mask? (1 = yes)
; ---- Input files ----
FR L
[ptcl_list]   ; particle list

FR L
[tilted_images]  ; tilted windowed image filename template

FR L
[cust_mas]cust_mas01   ; custom mask

; ---- Output files ----
FR L
[doc_dir]   ; doc file directory

FR L
[ali_dir]  ; ali ptcl directory

FR L
[dummy_rotate][doc_dir] ; tilted rotation doc filename template

FR L
[aligned_tilted][ali_dir]   ; align tilt image filename template

FR L
[shift_doc_tilt][doc_dir]   ; tilted shift doc filename template

FR L
[apsa_avg_tilt][ali_dir]   ; tilted intermediate avg filename
template

FR L
[alignment_doc][doc_dir]   ; tilt alignment doc filename template

FR L
[tilted_avg][ali_dir]   ; tilted centred average filename
template

FR L
[tilted_var][ali_dir]/vart   ; variance template for centred avg

FR L
[off_ali_doc][doc_dir]/offali   ; doc of not well aligning ptcls

FR L
[well_ali_doc][doc_dir]/wellali   ; doc of well aligning ptcls

; -- END BATCH HEADER --

;------------------------------------------------------------!
;   Create a dummy doc file for a rotational alignment where a
```

```
;   zero degree rotation angle would be found for all images

VM
echo "if(! -d [doc_dir]) mkdir -p [doc_dir]"|csh

VM
echo "if(! -d [ali_dir]) mkdir -p [ali_dir]"|csh


;GOTO LB8

; dummy variables for initial doc file
X31 = 0.0
x32 = 1.

; get #particles
UD N x62
[ptcl_list]

; get first ptcl#
UD 1,x61
[ptcl_list]
; get last ptcl#
UD x62,x63
[ptcl_list]

;   loop through particles
DO LB1 x11=1,X62
        ; get particle#
        UD IC,x11,x12
        [ptcl_list]

    SD X0,X12,X31,X32,X32
    [dummy_rotate]

IF (x95.EQ.1) GOTO lb2

;   Apply a circular mask on all the images
    MA
    [tilted_images]{*****X12}
    [aligned_tilted]{*****X12}
    x60      ; mask radius
    g        ; Gaussian cutoff
    a        ; background set to circumference average
    x51,x51  ; mask center
    (3.5)    ; Gaussian falloff

IF (x95.EQ.0) GOTO lb1

LB2

CP  ; copy images across to new aligned location
[tilted_images]{*****X12}
[aligned_tilted]{*****X12}

;   Apply a custom mask on all the images
    MM C
    [cust_mas]
```

```
    [aligned_tilted]{*****X12}

LB1

;   Alignment by translation on all the masked images
AP SA
[aligned_tilted]*****
[dummy_rotate]
(1)  ; group number from AP RA (set to dummy variable x32=1)
N    ; not checking 360-degree range
[shift_doc_tilt]001
[apsa_avg_tilt]001

;   Search the centre of gravity of the global average map

CG X21,X22
[apsa_avg_tilt]001
0.23                ; threshold value
x23 = x21
x24 = x22

;x25=x23-x51
;x26=x24-x51

 IF (x80.LT.1) GOTO LB5 ; avoids masking final particles

;   Apply X & Y translation (minus location of centre of gravity)
;   to all raw tilted-specimen images
DO LB4 X30 = 1,x62
    UD IC X30,X31,X32,X33,X34,X35,X36
    [shift_doc_tilt]001

IF (x92.LT.1) THEN
    X54=X34
    X55=X35
ELSE
    X54=X34-X23
    X55=X35-X24
ENDIF

    SD X0,X31,X52,X54,X55
    [alignment_doc]001

    X91 = X0

    RT SQ
    [tilted_images]{*****X31}
    scr001
    X52      ; rotation angle
    X54,X55  ; x-, y-shift

;   Apply a circular mask on the centred images
    MA
    scr001
    [aligned_tilted]{*****X31}
    x90
    g        ; Gaussian cut-off
    a        ; background set to circumference average
```

```
    x51,x51  ; mask centre
    (3.5)    ; Gaussian falloff

LB4

UD ICE
[shift_doc_tilt]001


;   Compute average and variance maps
AS R
[aligned_tilted]*****
[ptcl_list]
A
[tilted_avg]001
[tilted_var]001

; delete intermediate file
DE
scr001

LB8

; get #particles
UD N x62
[ptcl_list]

DO LB7 X30 = 1,x62

 UD IC X30,X31,X80,x81,x82
    [alignment_doc]001

IF (x81.GT.10) THEN
   SD x0,x31
   [off_ali_doc]
      ELSEIF (x81.LT.-10) THEN
       SD x0,x31
       [off_ali_doc]
      ELSEIF(x82.GT.10) THEN
            SD x0,x31
            [off_ali_doc]
      ELSEIF (x82.LT.-10) THEN
                 SD x0,x31
                 [off_ali_doc]
ENDIF

LB7

DOC SUBTRACT
[ptcl_list]
[off_ali_doc]
[well_ali_doc]
1

DOC REN
[off_ali_doc]
[off_ali_doc]
```

```
DOC REN
[well_ali_doc]
[well_ali_doc]

EN

LB5

;   Apply X & Y translation (minus location of centre of gravity)
;   to all raw tilted-specimen images
DO LB6 X30 = 1,x62
    UD IC X30,X31,X32,X33,X34,X35,X36
    [shift_doc_tilt]001

    ;X54=X34
    ;X55=X35

    X54 = X34-X23
    X55 = X35-X24

    SD X0,X31,X52,X54,X55
    [alignment_doc]001

    X91 = X0

    RT SQ
    [tilted_images]{*****X31}
    [aligned_tilted]{*****X31}
    X52      ; rotation angle
    X54,X55  ; x-, y-shift

LB6

UD ICE
[shift_doc_tilt]001


;   Compute average and variance maps
AS R
[aligned_tilted]*****
[ptcl_list]
A
[tilted_avg]001
[tilted_var]001

DO LB9 X30 = 1,x62

 UD IC X30,X31,X80,x81,x82
    [alignment_doc]001

IF (x81.GT.10) THEN
   SD x0,x31
   [off_ali_doc]
ELSEIF (x82.GT.10) THEN
   SD x0,x31
   [off_ali_doc]
ELSE
   SD x0,x31
```

```
   [well_ali_doc]
ENDIF

LB9


EN
```

**T3D_custom_mask.fed - Apply custom mask to particles**

```
; ---------------- Parameters ----------------
x45 = 0     ; initial average central? (1=yes)
x47 = 20    ; initial mask radius
x48 = 128   ; image dimensions
x49 = 84    ; x coordinate for initial mask
x50 = 76    ; y coordinate for initial mask
x51 = 0.01  ; threshold value
x52 = 0     ; Threshold ok? (1=yes)
x53 = 3     ; Filter type (3 = Gauss low-pass)
x54 = 0.06  ; Filter radius
x55 = 0.06  ; Final edge threshold
; ---------------- Input files ----------------
FR G
[avg_ptcl]  ; avg ptcl to mask

; ---------------- Output files ----------------
FR G
[mas_dir]mas   ; directory for mask

FR G
[cust_mas][mas_dir]/cust_mas01   ; custom mask

; ------------- END BATCH HEADER -------------

;---------------------------------------------------------------!
;   Apply a circular mask with Gaussian edge on all images
;---------------------------------------------------------------!

VM
echo "if(! -d [mas_dir]) mkdir -p [mas_dir]"|csh

DE
[mas_dir]/tmp_custom_mask01
DE
[mas_dir]/tmp_avg

IF (x45.EQ.1) GOTO LB1

MA  ; use off-centre initial mask (determine coords. using jweb
    ; marker option)
[avg_ptcl]
[mas_dir]/tmp_avg
x47 ; mask radius
g
e
0.0
```

```
x49,x50
(2.0)

GOTO LB2

LB1

x46 = x48/2+1 ; set mask centre to image centre

MA
[avg_ptcl]
[mas_dir]/tmp_avg
x47 ; mask radius
g
a
x46,x46
(3.5)

LB2

TH M
[mas_dir]/tmp_avg
[mas_dir]/tmp_custom_mask01
B   ; blank out below threshold value
x51 ; threshold value

IF (x52.eq.1)GOTO LB3 ; if threshold ok, continue, if not, end
ENDIF

EN

LB3

FQ NP
[mas_dir]/tmp_custom_mask01
[mas_dir]/tmp_custom_mask02
x53 ; filter type
x54 ; filter radius

TH M
[mas_dir]/tmp_custom_mask02
[cust_mas]
B
x55

EN D
```

**T3D_CC_TEST.spi - Cross-correlation test**

```
; ---------------- Parameters ----------------
x40 = 0.1 ; Percentage of ptcls to eliminate
```

```
x85 = 0   ; skip to manual threshold
x86 = 0.5 ; manual threshold input
x87 = 1   ; skip to manual CC ref input
x90 = 128 ; image dimension
x91 = 27  ; mask radius
x95 = 1   ; use custom mask? (1=yes)
; ---------------- Input files ----------------
FR G
[ptcl_list]  ; group-particle list

FR G
[filtered_dir] ; filtered directory

FR G
[filtered][filtered_dir] ; filtered particles template

FR G
[mask]   ; mask

FR G
[ali_ptcl] ; ali ptcl template

FR G
[ext_ref] ; external ref for CC if needed

; --------------- Output files ---------------
FR G
[ali_dir]  ; directory for aligned particles

FR G
[sum_dir][ali_dir]/sum   ; directory for sum images

FR G
[doc_dir][ali_dir]/doc   ; directory for doc files

FR G
[sum_image][sum_dir]/sum_img01   ; sum image

FR G
[var_image][sum_dir]/var_img01   ; variance image

FR G
[rot_avg_sum][sum_dir]/Ravg_img01   ; rotationally avg sum image

FR G
[cc_doc][doc_dir]/CCdoc01   ; CC doc

FR G
[cchist][doc_dir]/CChis01   ; Histogram of CC

FR G
[thresh][doc_dir]/CCthr01   ; CC thresholds

FR G
[seltotal][doc_dir]/seldoc1   ; ptcls over CC threshold

FR G
[regtotal][doc_dir]/rejdoc1   ; ptcls below CC threshold
```

```
; ------------- END BATCH HEADER -------------

MD
() OFF ; STOPS NEED FOR () AROUND NUMBER INPUTS IN DO LOOPS!!
SET MP

;-------------------------------------------------------------!
;   Create total sum of filtered images
;-------------------------------------------------------------!

VM
echo "Centering particles"; date

VM
echo "if(! -d [ali_dir]) mkdir -p [ali_dir]"|csh

VM
echo "if(! -d [doc_dir]) mkdir -p [doc_dir]"|csh

VM
echo "if(! -d [sum_dir]) mkdir -p [sum_dir]"|csh

IF (x95.EQ.1) GOTO LB8 ; skip to using custom mask
ENDIF

; compute a spherical mask
MO
[mask]
x90,x90
c
x91

LB8

IF (x85.EQ.1) GOTO LB6 ; skip to manual threshold
ENDIF

IF (x87.EQ.1) GOTO LB7 ; skip to manual CC reference input
ENDIF

AS R
[filtered]***** ; template of images to be averaged
[ptcl_list]
A ; all images will be used to create average
tmp_sum_img
tmp_var_img

CG x10 x11
tmp_sum_img
.23
SH
tmp_sum_img
[sum_image]
-x10,-x11

CG x10 x11
tmp_var_img
```

```
.23
SH
tmp_var_img
[var_image]
-x10,-x11

; remove temporary files
DE
tmp_sum_img
DE
tmp_var_img

;-------------------------------------------------------------!
;   Rotationally average total sum to create initial reference
;-------------------------------------------------------------!

RO I
[sum_image]
[rot_avg_sum]

LB7

IF (x87.eq.1) THEN       ; set manual CC reference input
[rot_avg_sum]=[ext_ref]  ; overrides rotational average
ENDIF

; get #particles
UD N x62
[ptcl_list]

;   loop through particles
DO LB1 x11 = 1,X62
        ; get particle#
        UD IC,x11,x12
        [ptcl_list]

CC C x20,x21,x22,x23,x24,x25
[ali_ptcl]{*****x12}
[rot_avg_sum]
[mask]

SD x11,x12,x20,x21,x22,x23,x24,x25
[cc_doc]

LB1

UD ICE
[ptcl_list]

;;;; MAKE HISTOGRAM DOCUMENT FILE ;;;;

HD D
[cc_doc]
(2)
(128)
[cchist]

;;;; THRESHOLD THE PARTICLES TO REMOVE ANOMALOUS PARTICLES ;;;;
```

```
x77 = 1 ; COUNT, NEED TO WORK THIS IN TO GO THRO CLASSES

SD /    CC_THRESH   N_ABOVE_THR   N_BELOW_THR   N_TOTAL
[thresh]

; Find the total number of particles (x30)
   x30 = 0

   UD N,x28
   [cchist]

   DO LB2 x31 = 1,x28
      UD x31,x32,x33
      [cchist]

      x30 = x30 + x33
   LB2

   x35 = x40 * x30        ; Number of particles to eliminate

   x50 = 0                ; Cumulative no. of particles
   x51 = 0

   DO LB4 x31 = 1,x28
      UD x31,x32,x33
      [cchist]
      x50 = x50 + x33

      IF (x50.GT.x35) THEN
         x31 = x31-1

         UD x31,x32,x33    ; Get the previous bin
         [cchist]

         GOTO LB3
      ENDIF

      x51 = x50   ; Cumulative no. of particles from previous bin

   LB4

   LB3

   x42 = x30 - x51 ; N_above = total - N_below

;Save CC_threshold, N_above_threshold, N_below_threshold, total_N

   SD x77,x32,x42,x51,x30
   [thresh]

UD E

LB6

x61 = 0
x63 = 0
```

```
; get #particles
UD N x62
[cc_doc]

IF (x85.EQ.1) THEN
    x32=x86
ELSE
    UD x77,x32
    [thresh]
ENDIF

;   loop through particles
DO LB5 x11=1,x62
        ; get particle# + CC
        UD IC,x11,x12,x82
        [cc_doc]

      IF (x82.GE.x32)THEN
       x61 = x61+1

       SD x61,x12,x82
       [seltotal]
   ENDIF
        IF (x82.LT.x32) THEN
       x63 = x63+1

       SD x63,x12,x82
       [regtotal]
   ENDIF
LB5

EN D
```

**T3D_APSR.cor - Reference-free rotational and translational alignment**

```
; RUN REFERENCE-FREE ALIGNMENT
;
; NOTE: requires counter.cor

; ---------------- Parameters ----------------
x10 = 1         ; number of alignment runs
x55 = 2         ; first ring radius for alignment
x50 = 128       ; image dimension
x51 = x50/2+1   ; image centre coordinates
x60 = 28        ; particle diameter
x65 = 25        ; radius for mask
x70 = 1         ; mask final particles (1=NO)
; ---------------- Input files ----------------
FR G
[group_particles]  ; group-particle list

FR G
[flt_dir] ; filtered directory

FR G
[flt_ptcls][flt_dir]/NCF_unt*****   ; filtered particles
```

```
FR G
[centering_ref] ; centring ref image (use * as default)

; --------------- Output files ---------------
FR G
[masked_dir]mas   ; directory for masked particles

FR G
[masked_ptcls][masked_dir]/mas*****   ; masked images

FR G
[apsr_dir]                    ; output AP SR directory

FR G
[apsr_avg][apsr_dir]/{**x19}avg***  ; AP SR average

FR G
[final_avg][apsr_dir]/final_avg     ; final average image

FR G
[final_var][apsr_dir]/final_var       ; final varinace image

FR G
[apsr_doc][apsr_dir]/{**x19}temp***   ; AP SR doc file

FR G
[doc_dir]  ; Document directory

FR G
[final_apsr_doc][doc_dir]/apsr_doc01   ; Final APSR document file

FR G
[apsr_counter][apsr_dir]/rd{**x10}count ; AP SR iteration-counter

FR G
[aligned][apsr_dir]/apsr*****   ; aligned particles

FR G
[aligned_m][apsr_dir]/apsr_m*****   ; aligned and masked ptcls

; ------------- END BATCH HEADER -------------

; calculate radius for last alignment ring
x30 = x60/2

VM
echo "Aligning particles"; date
VM
echo "if(! -d [masked_dir]) mkdir -p [masked_dir]"|csh
VM
echo "if(! -d [apsr_dir]) mkdir -p [apsr_dir]"|csh
VM
echo "if(! -d [doc_dir]) mkdir -p [doc_dir]"|csh

; get #particles
UD N x11
[group_particles]
```

```
; loop through particles
    DO LB1 x12=1,x11
        ; get particle#
        UD IC,x12,x13
        [group_particles]

;   apply a circular mask on the raw images and stack them
    MA
    [flt_ptcls]x13
    [masked_ptcls]x13
    x65     ; mask radius
    g        ; Gaussian cut-off
    a        ; changed this from C to external
    x51,x51  ; mask centre
    (6.0)    ; Gaussian falloff
LB1

UD ICE
[group_particles]

; loop through attempts
DO LB2 x19 = 1,x10
    AP SR
    [masked_ptcls]     ; particles to be aligned
    [group_particles]  ; selection file
    x60                ; expected size of the object
    x55,x30            ; first and last ring radius
    [centring_ref]     ; centring image
    [apsr_avg]
    [apsr_doc]

    ; count number of apsr docs
    @counter
    100  ; maximum #iterations

    ; get #iterations
    UD N, x88
    [apsr_counter]

    ; loop through particles
    DO LB3 x12 = 1,x11
        ; get particle#
        UD IC,x12,x13
        [group_particles]

        ; get alignment parameters
        UD IC,x12,x21,x22,x23
        [apsr_doc]x88

         x41 = x22
         x42 = x23

      SD x12,x13,x21,x41,x42
      [final_apsr_doc]

        RT SQ
        [flt_ptcls]x13
```

```
         [aligned]x13 ; previously tmp_alignedx13
         (x21,1)  ; angle, scale
         x41,x42  ; x,y-shift

        IF (x70.EQ.1) GOTO lb3

        ;apply a circular mask on the aligned images
         MA
         [aligned]x13
         [aligned_m]x13
         x65      ; mask radius
         g        ; Gaussian cut-off
         a        ;
         x51,x51  ; mask centre
         (6.0)    ; Gaussian falloff
    LB3
    ; end particle-loop

    UD ICE
    [apsr_doc]x88


LB2
; end attempt-loop

IF (x70.EQ.1) GOTO lb5

;   Compute average and variance maps
AS R
[aligned_m]
[group_particles]
A
[final_avg]
[final_var]

GOTO LB6

LB5

;   Compute average and variance maps
AS R
[aligned]
[group_particles]
A
[final_avg]
[final_var]

LB6

; clean up loop

; loop through particles
    DO LB4 x12 = 1,x11
        ; get particle#
        UD IC,x12,x13
        [group_particles]

DE A
```

```
[apsr_dir]/01temp001
IF (x70.EQ.1) GOTO lb7
DE
[aligned]x13
LB7
DE
[masked_ptcls]x13

LB4

UD ICE  ; close doc
[group_particles]

VM
echo "Done"; date

EN
```

**T3D_MSA.fed - Multivariate statistical analysis**

```
MD
TR ON

; ----- Parameters -----
x25 = 2.0   ; additive constant, in case there are pixel values<0
[img_dim] = 128    ; image dimension
[mask_rad] = 25    ; mask radius
x20 = 20    ; number of eigenvectors to compute
x30 = 0    ; Use custom mask? (1=yes)

; ---- Input files ----
FR L
[ptcl_list]  ; particle list document

FR L
[aligned_ptcls]  ; aligned image filename template

; ---- Output files ----
FR G
[mask_dir]mas   ; directory for mask

FR L
[final_mask][mask_dir]/mas01   ; mask to apply to particles

FR G
[coran_dir]coran ; directory for correspondence analysis

FR L
[coran][coran_dir]/coran   ; prefix for correspondence analysis
;                                  results files :
;                                  coran1_SEQ.hbl
;                                  coran1_IMC.hbl
;                                  coran1_PIX.hbl
;                                  coran1_EIG.hbl
```

```
FR L
[ps_map][coran_dir]/map.ps   ; postscript eigenvalue map

; -------------------- END BATCH HEADER --------------------

VM
echo "if(! -d [mask_dir]) mkdir -p [mask_dir]"|csh
VM
echo "if(! -d [coran_dir]) mkdir -p [coran_dir]"|csh

IF (x30.EQ.1) GOTO LB1

; compute a spherical mask :
MO
[final_mask]
[img_dim],[img_dim]
c
[mask_rad]

LB1

;   compute the contingency table (or sequential file) SEQ001.hbl
:
CA S
[aligned_ptcls]
[ptcl_list]
[final_mask]
x20
c
x25    ; additive constant*
[coran]
;
; Computes projection map corresponding to axes 1 and 2
CA SM  ; correspondance analysis, show map and eigenvalues
i  ; each image is represented by a vector in n-dimensional space
[coran] ; coordinate file prefix
1       ; number of horizontal patches
1,2     ; eigenvectors to map
s       ; represent by a symbol
+       ; symbol
Y       ; prepare postscript map
3.0     ; range, measured in standard deviations

[ps_map]
12,12
0



0



; The 8 blank lines are used in CA SME to accept default values.

EN
```

**T3D_Imp+Rec.fed - Generate importance and reconstituted images**

```
MD
TR ON

; ----- Parameters -----
x50 = 20    ; number of eigenvectors to compute

; ---- Input files ----
FR L
[coran]coran/coran   ; prefix for correspondence analysis
;                                    results files :
;                                    corantest1_SEQ.hbl
;                                    corantest1_IMC.hbl
;                                    corantest1_PIX.hbl
;                                    corantest1_EIG.hbl

FR G
[particle_list]   ; particle list

; ---- Output files ----
FR L
[imp_dir]Imp_img ; importance image directory

FR L
[importance_montage][imp_dir] ; montage of importance images

FR L
[recon_montage][imp_dir] ; montage of reconstituted images

; --------------- END BATCH HEADER ----------------

VM
echo "if(! -d [imp_dir]) mkdir -p [imp_dir]"|csh

; TEMPORARY FILES:
FR L
[imp_img]Imp_img/imp   ; imp image filename temp

FR L
[recon_img]Imp_img/rec  ; reconstituted image filename template

x11 = 100. ; initialize importance image file # for negative-most
           ; eigenvalue
x21 = 100. ; initialize reconstituted image file # for negative-
           ;most eigenvalue

; loop through eigenvectors
DO LB1 x10 = 1,x50
;   importance images of (-) & (+) extremities along axes
    x11 = x11+1.  ; increment file # for negative-most eigenvalue
    x12 = x11+1.  ; increment file # for positive-most eigenvalue

    SD C
    [coran]
    x10         ; factorial axis to be used
    tempdoc1
```

```
;
;get #particles
    UD N x61
[particle_list]

; loop through particles
DO LB2 x62 = 1, x61
    ; get particle#
    UD IC x62,x63
    [particle_list]
    UD S x63,x73
    tempdoc1
    SD x62,x63,x73
    tempdoc2
LB2
UD ICE
[particle_list]
;ud e
;tempdoc1
SD S
tempdoc2
1,x61
(2)
(1) ; SHOULD INPUT ORIGINAL PARTICLE NO. IN THE 3RD COLUMN OF DOC
tempdoc{***x10}

UD 1,x64,x65,x66
tempdoc{***x10}


    CA SRD
    [coran]       ; correspondence analysis results files prefix
    x66                   ; file number (ORIGINAL PARTICLE NO.)
    x10                   ; factorial axis to be used
    tmp****

    CP
    tmp{****x64}
    [imp_img]{***x11}    ; output importance image


    UD x61,x67,x68,x69
    tempdoc{***x10}

    CA SRD
    [coran]        ; correspondence analysis results files prefix
    x69                   ; file number
    x10                   ; factorial axis to be used
    tmp****

    CP
    tmp{****x67}
    [imp_img]{***x12}    ; output importance image

    SD E
    tempdoc2

    ; clean-up temporary files
```

```
    DE
    tempdoc1
    DE
    tempdoc2
    DE
    tmp{****x64}
    DE
    tmp{****x67}



    x11 = x12

;   reconstituted images of (-) & (+) extremities along axes
    x21 = x21+1.  ; increment file number for negative-most
                  ; eigenvalue
    x22 = x21+1.  ; increment file number for positive-most
                  ; eigenvalue

    CA SRA
    [coran]   ; correspondence analysis results files prefix
    x10                     ; factorial axis to be used
    (-0.2)                  ; eigenvalue for second file
    [recon_img]{***x21}     ; output importance image


    CA SRA
    [coran]     ; correspondence analysis results files prefix
    x10                    ; factorial axis to be used
    (0.2)                  ; eigenvalue for second file
    [recon_img]{***x22}    ; output importance image

    x21 = x22

LB1

;   montages of importance and reconstitution images
MN S
[imp_img]***
102,104,106,108,110,112,114,116,118,120,122,124,126,128,130,132,
134,136,138,140,101,103,105,107,109,111,113,115,117,119,121,123,
125,127,129,131,133,135,137,139
x50,2  ; images per row, margin width
2    ; margin value
[importance_montage]

; ,122,124,126,128,130,132,134,136,138,140
; ,121,123,125,127,129,131,133,135,137,139

MN S
[recon_img]***
102,104,106,108,110,112,114,116,118,120,122,124,126,128,130,132,1
34,136,138,140,101,103,105,107,109,111,113,115,117,119,121,123,12
5,127,129,131,133,135,137,139
x50,2  ; images per row, margin width
2    ; margin value
[recon_montage]
```

```
;   delete intermediate files :
DE A
[imp_img]101
DE A
[recon_img]101
DE A
tempdoc001

VM
echo "Done"; date

EN D
```

**T3D_hac.fed - Hierarchical ascendant classification**

```
MD
TR ON
MD
TERM ON ; prints output to the terminal

; ----- Parameters -----
x10 = 0 ; make dendrodoc = (0) skip and run classification = (1)

FR L
[eig]1     ; eigenvectors to use

x13 = 0    ; eigenvector weighting ('0'-->all 1's)
x14 = 5    ; clustering criterion: Ward's method
x26 = 382  ; maximum dendrogram branch point
x27 = 60   ; cut-off for branch points

; ---- Input files ----
FR L
[image_coords]coran/coran_IMC ; image coords from MSA by CA S

FR L
[dendro_doc]classes/dhac001   ; dendrogram doc file

FR L
[aligned_untilted] ; untilted aligned image filename template

; ---- Output files ----
FR L
[hac_dir]classes   ; HAC directory

FR L
[hac_img][hac_dir]/img   ; HAC image directory

FR L
[hac_doc][hac_dir]/doc   ; HAC document file directory

FR L
[ps_dendro][hac_img]   ; PostScript, truncated dendrogram plot

FR L
```

```
[class_count_doc][hac_doc]/dcount01   ; # images per class doc

FR L
[class_select_doc][hac_doc]/dcla   ; unsorted class selection doc
                                   ; filename template
FR L
[class_doc][hac_doc]/cldoc   ; sorted class selection doc

FR L
[class_avg][hac_img]/clavg   ; class average filename template

FR L
[class_var][hac_img]/clvar   ; class variance filename template

FR L
[class_stats][hac_doc]/sdvar ; class variance stats

FR L
[class_avg_montage][hac_img]/mclav001   ; class average montage

FR L
[class_var_montage][hac_img]/mclva001   ; class variance montage

; ---------------- END BATCH HEADER -----------------

VM
echo "if(! -d [hac_dir]) mkdir -p [hac_dir]"|csh

VM
echo "if(! -d [hac_img) mkdir -p [hac_img]"|csh

VM
echo "if(! -d [hac_doc) mkdir -p [hac_doc]"|csh

IF(x10.EQ.1.0)GOTO LB19

;   Runs classification and creates dendrogram doc file
CL HC
[image_coords]
[eig]  ; eigenvectors to use
x13  ; eigenvector weighting ('0'-->all 1's)
x14  ; clustering criterion: Ward's method
*    ; postscript file? (make a truncated one later)
y    ; dendrogram doc file?
[dendro_doc]

EN
 (Check dendrogram shape and note Max and Min values)

LB19
MD
TERM OFF ; no output to terminal

;   Compute the number of images per class
x11 = (x27/x26)*100  ; branch cut-off divided by max branch point

;   Starts again HAC and creates dendrogram file
CL HC
```

```
[image_coords]
[eig]  ; eigenvectors to use
x13  ; eigenvector weighting ('0'-->all 1's)
x14  ; clustering criterion: Ward's method
t    ; truncated postscript file
x11  ; cut-off (Scale value) CHANGED FROM x27
[ps_dendro]
n

;!!! MOVED TO ABOVE  Compute the number of images per class !!!
; x11=(x27/x26)*100  ; branch cut-off divided by max branch point

CL HD  ; hierarchical classification, calculate classes
x11
[dendro_doc]
[class_count_doc]

;   Creating doc files containing the image list for each class
CL HE
x11
[dendro_doc]
[class_select_doc]***

UD N x28
[class_count_doc]

; sort class doc file by particle number
DO LB1 x29 = 1,x28

      DOC SORT
      [class_select_doc]{***x29}
      [class_doc]{***x29}
      (1)
      Y

      DE
      [class_select_doc]{***x29}

;   Compute average and variance maps for each class
    AS DC
    [aligned_untilt]*****
    [class_doc]{***x29}
    a  ; sum all images
    [class_avg]{***x29}
    [class_var]{***x29}

; get SDVAR
        FS x99,x99,x99,x15
        [class_var]{***x29}
; get class-size
        UD N,x16
        [class_doc]{***x29}

        ; write to doc
        SD x29,x29, x15,x16
        [class_stats]_t
        ; x15 = sdvar, x16 = class-size
```

```
LB1

 ; sort
    DOC SORT
    [class_stats]_t
    [class_stats]
    (2)  ; column# to sort: SDVAR

    DE
    [class_stats]_t
;
;   Creates montage files of average and variance maps
MN S
[class_avg]***
1-x28 ; file numbers
4,2   ; images per row, margin width
2     ; margin value
[class_avg_montage]

MN S
[class_var]***
1-x28 ; file numbers
4,2   ; images per row, margin width
2     ; margin value
[class_var_montage]

EN
```

**T3D_resclac.cor - Calculate 2D class nominal resolution**

```
; CALCULATE RESOLUTION FOR EACH CLASS
;
; Requires the following subroutines:
;    frcmask.cor
;    frcrescalc.cor

; ------------- Parameters -------------
x28 = 2.62    ; pixel size (in Angstroms)
x75 = 1        ; number of groups
x56 = 1    ; progress-message displayed every Nth class

; ------------- Input files -------------
;fr l
;[parameter_doc]   ; parameter doc
;fr g
;[group_dir]   ; input group directory
;fr g
;[coran_dir][group_dir]   ; input Coran directory
fr g
[class_dir]  input class directory
fr l
[list_classes][class_dir]/dcount01   ; list of classes
fr g
[class_doc][class_dir]/cldoc***   ; class doc
fr g
```

```
[align_dir]   ; input aligned directory
fr g
[aligned][align_dir]/apsr*****   ; aligned particles
fr g
[mask]mas/mask01   ; mask

; ------------ Output files ------------
fr g
[final_resolution][class_dir]/frc_doc   ; final FRC doc

; ---------- END BATCH HEADER ----------

vm
echo "Calculating 2D resolution"; date

fr g
[group_frc][class_dir]/rfdoc{**x11}
fr g
[group_res][class_dir]/resdoc{**x11}

; get pixel size
;ud 5,x28
;[parameter_doc]
;ud e

; get #classes (x82)
ud n,x82
[list_classes]

; loop through groups
do lb1 x77=1,x75
    de
    [final_resolution]
    SD /      CLASS        0.5FRC       3-SIGMA
    [final_resolution]

    ; loop through classes
    do lb2 x11=1,x82
        if (int(x11/x56).eq.x11/x56) then
            vm
            echo "Working on class #{***x11} out of {***x82}"
        endif

        ; calculate Fourier ring correlations
        @frcmask
        [class_doc]x11  ; INPUT
        [aligned]       ; INPUT
        [mask]          ; INPUT
        [group_frc]     ; OUTPUT

        ; calculate nominal resolution
        @frcrescalc(x28,x11)
        [group_frc]  ; INPUT
        [group_res]  ; OUTPUT

        ; get 0.5-frc (x13), 3-sigma resolutions (x14)
        ud 1,x13,x14
        [group_res]
```

```
        ud e

        ; if no 0.5-frc resolution, then skip ahead
        if(x13.eq.0)then
            x15=0
            goto lb3
        endif

        ; convert from 1/A to Angstroms
        x15=1/x13

        lb3  ; jump here if no 0.5-frc resolution

        ; if no 3-sigma resolution, then skip ahead
        if(x14.eq.0)then
            x18=0
            goto lb4
        endif

        ; convert from 1/A to Angstroms
        x18=1/x14

        lb4  ; jump here if no 3-sigma resolution

        ; write resolution to doc file
        sd x11,x11,x15,x18
        [final_resolution]
        ; x15==0.5-resolution, x18==3-sigma

        ; clean up
        de
        [group_frc]
        de
        [group_res]
    lb2
    ; end class-loop

    sd e
    [final_resolution]
lb1
; end group-loop

vm
echo "Done"; date

en d
```

**T3D_Inter-class_ptcl_tracker.spi - Track movement of particles between classes**

```
; ---------------- Parameters ----------------

x10 = 4    ; number of classes
```

```
; ---------------- Input files ----------------
FR L
[doc_dir]classes ; directory for doc files

FR L
[APMS_doc][doc_dir]/APMSdoc001 ; APMS doc file

FR L
[doc_old][doc_dir]/cldoc  ; tempate for old doc files

FR L
[doc_new][doc_dir]/APMS_cldoc  ; tempate for new doc files

; ---------------- Output files ----------------
FR L
[doc_moves][doc_dir]/mv_cldoc ; doc file of ptcl movements

; ------------- END BATCH HEADER -------------

UD N x21
[APMS_doc]

DO LB1 x30 = 1,x10

      DO LB2 x20 = 1,x21

      UD IC x20,x22,x23,x24,x25,x26,x27
      [APMS_doc]

      IF (x22.EQ.x30) THEN
      SD x20,x27
      [doc_dir]/temp{***x30}
      ENDIF

      LB2

DOC REN
[doc_dir]/temp{***x30}
[doc_new]{***x30}

DE
[doc_dir]/temp{***x30}

LB1

UD ICE
[APMS_doc]

EN D

DO LB3 x11 = 1,x10

UD N x12
[doc_old]{***x11}

      DO LB4 x13 = 1,x12
```

```
      UD IC x13,x14
      [doc_old]{***x11}

            DO LB5 x15 = 1,x10

            UD FIND x0,x16
            [doc_new]{***x15}
            (1),x14

            IF (x0.EQ.0) GOTO LB7

            IF (x0.NE.0) GOTO LB6

            LB7
            UD ICE
            [doc_new]{***x15}
            LB5

      LB6

      SD x0,x14,x15
      [doc_moves]{***x11}

      UD ICE
      [doc_new]{***x15}

      LB4

UD ICE
[doc_old]{***x11}

LB3

EN
```

**T3D_Euler_doc.fed - Create Euler angle document file**

```
; ----- Parameters -----
x40 = 124.  ; index number for relevant line in TILT_PARAMS_DOC
X61 = 22    ; Number of tilt pairs
x12 = 1     ; set to 1 to skip making all_tilt doc file

; ---- Input files ----
FR L [tilt_params_doc]P_{**x18}/dcb  ; template for doc files
                                     ; containing tilt parameters
FR L
[align_doc_untilt]apsr_doc01 ; untilted alignment doc filename
                             ; template

FR L
[all_ptcl_list]listparticles   ; list of all particles

FR L
[ptcl_by_micro_list]   ; particle-by-micrograph list template

FR L
```

```
[class_ptcl_list]   ; list of particles in class

FR L
[Tptcl_cntr_doc]*   ; tilted centring doc

FR L
[Uptcl_cntr_doc]*   ; untilted centring doc

; ---- Output files ----
FR L
[all_tilt_doc]all_tilt ; doc with tilt angles for all particles

FR L
[3D_dir]r3d ; 3D directory

FR L
[angles_doc][3D_dir]dang   ; Eulerian angle doc file template

; ------------------ END BATCH HEADER ----------------

MD
TR OFF


VM
echo "if(! -d [3D_dir]) mkdir -p [3D_dir]"|csh

IF(x12.EQ.1) GOTO LB5 ; SKIP MAKING ALL_TILT DOC FILE.

DO LB1 x18 = 1,x61

UD N x50
[ptcl_by_micro_list]_{***x18}
UD x50,x51
[ptcl_by_micro_list]_{***x18}
UD x40,x31,X44,x36
[tilt_params_doc]{***x18}
UD 1,x62
[ptcl_by_micro_list]_{***x18}

  DO LB2 X10 = x62,x51


  SD x10,x44,x31,x36 ; SAVED AS: PSI, THETA, PHI
  [all_tilt_doc]

  IF (x10.EQ.x51) GOTO LB1

  LB2

LB1

UD E

UD ICE
[all_ptcl_list]
```

```
UD N x70
[class_ptcl_list]

       DO LB3 X71 = 1,x70

       UD IC x71,x72
       [class_ptcl_list]

       UD x72,x31,x32,x34 ; TRY USING TILTED PHI AS UNTILTED PSI
          [all_tilt_doc]

       UD FIND x0,x0,x33
       [align_doc_untilt]
          (1),x72

  x33 = -x33     ; correct for sign convention
  x33 = x33 - x31  ; subtract relative rotation between tilt pair
  x34 = 0.0 ; dummy variable for Eulerian angle psi

; X31 IS CORRECTED TILTED PTCL PSI ANGLE. THIS IS SUBTRACTED FROM
; UNTILTED PTCL PHI ANGLE.

;Stores the three eulerian angles PSI, THETA, & PHI in a doc file
; PSI IS PHI from tilted ptcls, THETA IS TILT FROM TILTED PTCLS,
; PHI IS TOTAL CORRECTED ROTATION with tilted gamma subtracted

       SD x72,x34,x32,x33 ; PSI,THETA,PHI
       [angles_doc]001

       LB3

;-------------------------------------------------------------!
; For clarity, we add or subtract 360 to angle phi so that its
; value is restricted between 0 ET 360 (it helps if you want to
; check angles)
;-------------------------------------------------------------!

; loop through
DO LB9 X80 = 1,4
   X81 = X80+1.  ; counter for new ANGLES_DOC

UD N x70
[class_ptcl_list]

;  loop through particles
   DO LB8 X10 = 1,x70
      UD x10,x72
      [class_ptcl_list]
;     get old angles
      UD S x72,X21,X22,X23
      [angles_doc]{***x80}

;     force phi to be between 0 and 360

      IF(X23.GT.0.0)GOTO LB20

     X23 = X23+360.
```

```
      LB20

      IF(X23.LT.360)GOTO LB22

      X23=X23-360.

      LB22

      SD x72,X21,X22,X23
      [angles_doc]{***x81}
   LB8
LB9

EN
```

**T3D_Initial_3D.fed - Generate 3D model**

```
MD
TR ON
MD
TERM ON

; ---------------- Parameters ----------------

x10 = 22    ; Particle radius
x11 = 5.0E-5    ; Correction weight
x12 = 100    ; iteration limit
x13 = 0    ; mode (0 = no constraint 1 = smoothing)
x14 = 0.990 ; smoothing constant
x15 = 2    ; 0=Use BP RP   1=Use BP 32F   2=Use BP CG
; ---- Input files ----
FR L [aligned_tilted]; tilted aligned tilted image filename
template

FR L
[class_select_doc] ; class selection doc filename template

FR L
[angles_doc]; Eulerian angle doc file template

FR L
[ang_sym]*   ; symmetry doc file (leave blank if no symmetry)

FR L
[ran_doc1]; odd image file

FR L
[ran_doc2]; eve image file


; ---- Output files ----
FR G
[3d_dir]; 3d directory

FR L
```

```
[class_vol][3d_dir]/model001   ; 3d model

; --------------- END BATCH HEADER ----------------

VM
echo "if(! -d [3d_dir]) mkdir -p [3d_dir]"|csh

IF (x10.EQ.-1) THEN
   x10 = INT( (0.95*128)/2.0 )
ENDIF

IF (x15.EQ.1) GOTO LB1

IF (x15.EQ.2) GOTO LB2

; 3D Reconstruction of image class No.1 :

BP RP
[aligned_tilted]*****
[class_select_doc]
x10        ; particle radius
[angles_doc]
[ang_sym]
[class_vol]
x11,0     ; correction weight, correction limit
x12,x13   ; iteration limit, mode 1 (smoothing)
1.0,1.5   ; minimum, maximum values in projections
x14       ; smoothing constant


; 3D Reconstruction of image class No.1 :

BP RP
[aligned_tilted]*****
[ran_doc1]
x10        ; particle radius
[angles_doc]
[ang_sym]
[class_vol]_odd
x11,0    ; correction weight, correction limit
x12,x13  ; iteration limit, mode 1 (smoothing)
1.0,1.5  ; minimum, maximum values in projections
x14      ; smoothing constant

; 3D Reconstruction of image class No.1 :

BP RP
[aligned_tilted]*****
[ran_doc2]
x10        ; particle radius
[angles_doc]
[ang_sym]
[class_vol]_eve
x11,0     ; correction weight, correction limit
x12,x13   ; iteration limit, mode 1 (smoothing)
1.0,1.5   ; minimum, maximum values in projections
x14       ; smoothing constant
```

```
EN

LB1

BP 32F
[aligned_tilted]*****
[class_select_doc]
[angles_doc]
[ang_sym]
[class_vol]
[class_vol]_odd
[class_vol]_eve

EN

LB2

BP CG3
[aligned_tilted]*****
[class_select_doc]
x10
[angles_doc]
F
[class_vol]
;[class_vol]_odd
;[class_vol]_eve
5.0E-5,0.0
25,1
200

EN

BP CG
[aligned_tilted]*****
[ran_doc1]
x10
[angles_doc]
*
[class_vol]_odd
1.0E-5,0.0
25,1
1000

BP CG
[aligned_tilted]*****
[ran_doc2]
x10
[angles_doc]
*
[class_vol]_eve
1.0E-5,0.0
25,1
1000

EN
```

**T3D_Refine_3D.fed - Refine initial 3D model by projection matching**

```
 ; ---------------- Parameters ----------------
x81 = 25   ; No. of iterations
x95 = 0    ; Force iterations? (1=yes)
x10 = 3    ; Translational search range (pixels)
x11 = 1    ; Tran. step size (pixels)
x12 = 2    ; First ring for Rot. align (pixels)
x13 = 17   ; Last ring for Rot. align (pixels)
x14 = 1    ; Ring step
x15 = 1    ; Ray step
x16 = 0.0  ; Range of angular. search (deg): 0 = no restriction
x17 = 0.0  ; Angle change threshold (deg)
x18 = 0    ; Check mirrored positions (0 = NOCHECK / 1 = CHECK)
x19 = 20   ; radius of object (for projections)
x90 = 500  ; initial CCROT cut-off for ptcls to use in 3D model
x91 = 0    ; step increase in CCROT per iteration
x20 = 50   ; radius for final 3D (use -1 for 95% winsize / 2)
x65 = 2.62 ; Pixel size
x40 = 0.5  ; Cut-off for Fourier shell correlation curve
x41 = 1    ; Use the Nth time the curve crosses the threshold
x99 = 50   ; Tilt angle (deg)

; ---- Input files ----
FR L
[prj_doc]prj_doc001   ; Doc file for the list of projections

FR L
[prj_ang]prj_ang001   ; Doc file for the stored projection angles

FR L
[cntr_tilt_ptcls]sali  ; Template for centered tilted ptcls

FR L
[ptcl_list]            ; List of particles to use in realignment

FR L
[tilt_ptcl_ang]   ; List of centred tilted ptcl. angles

FR L
[3D_model]   ; model template

; ---- Output files ----
FR L
[3D_dir]  ; output dir

FR L
[PJ_ptcls][3D_dir]/proj@****   ; Template for particle
reprojections

FR L
[APSH_doc][3D_dir]/APSHdoc   ; Output APSH document file

FR L
[recntr_dir][3D_dir]/ali   ; re-ali directory
```

```
FR L
[recntr_tilt_ptcls][recntr_dir]/re-ali   ; Template for refined
shift aligned tilted ptcls

FR L
[select_doc][3D_dir]/sel_doc   ; select doc

FR L
[FSC_doc][3D_dir]/FSCdoc   ; FSC doc file

FR L
[res][3D_dir]/RESdoc   ; Doc file with resolution data

; --------------- END BATCH HEADER ----------------

VM
echo "if(! -d [3D_dir]) mkdir -p [3D_dir]"|csh

VM
echo "if(! -d [recntr_dir]) mkdir -p [recntr_dir]"|csh

MD
TR OFF

IF (x20.eq.-1) THEN
x20 = INT( (0.95*128)/2.0 )
ENDIF

DO LB1 x80 = 1,x81

x82 = x80+1

VM
rm [3D_model]{***x80}_odd.spi

VM
rm [3D_model]{***x80}_eve.spi


VO EA
1
45,50
0,359.9
[prj_ang]


PJ 3Q
[3D_model]{***x80}
x19
[prj_doc]
[prj_ang]
[PJ_ptcls]


AP SH
[PJ_ptcls]
[prj_doc]
```

```
x10,x11
x12,x13,x14,x15
[prj_ang]
[cntr_tilt_ptcls]*****
[ptcl_list]
[tilt_ptcl_ang]
x16,x17
x18
[APSH_doc]{***x80}


;get #particles

UD N x61
[ptcl_list]

        ; loop through particles

          DO LB2 x62=1,x61

        ; get particle#

            UD IC x62,x63
            [ptcl_list]

          UD x63,x22,x23,x24,x25,x26,x27,x28,x29,x30,x31,x32
          [APSH_doc]{***x80}

          IF (x32.GT.x90) THEN
          SD x0,x63
          [select_doc]{***x80}
          ENDIF

; APPLY NEW X,Y SHIFTS + ROTATIONS TO TILTED PTCLS FROM APSH

RT SQ
[cntr_tilt_ptcls]{*****x63}
[recntr_dir]/{***x80}re-ali{*****x63}
x27
x28,x29

        LB2

        DOC REN
        [select_doc]{***x80}
        [select_doc]{***x80}

        x90 = x90+x91

        UD ICE
        [ptcl_list]

        BP 32F
        [recntr_dir]/{***x80}re-ali*****
        [select_doc]{***x80}
        [APSH_doc]{***x80}
        *
        [3D_model]{***x82}tmp
```

```
         [3D_model]{***x82}_odd
         [3D_model]{***x82}_eve

      CG [xcg],[ycg],[zcg]
      [3D_model]{***x82}tmp

      SH F
      [3D_model]{***x82}tmp
         [3D_model]{***x82}
         -[xcg],-[ycg],-[zcg]

      DE
      [3D_model]{***x82}tmp

         ; Create FSC curve

         RF 3
         [3D_model]{***x82}_odd
         [3D_model]{***x82}_eve
         (0.5)
         (0.2,2.0)
         C
         x99
         (3)
         [FSC_doc]{***x82}

         DOC REN
         [FSC_doc]{***x82}
         [FSC_doc]{***x82}

; Calculate resolution

x42 = 1 ; counter to compare with x41

UD N, x51
[FSC_doc]{***x82}

; Get the first line (x21 = previous, x31 = current)
UD 1,x21,x22,x23
[FSC_doc]{***x82}

         DO LB3 x50 = 1,x51

         UD x50,x31,x32,x33
         [FSC_doc]{***x82}

         IF (x33.LT.x40) THEN
            IF (x23.GT.x40) THEN
               IF (x42.EQ.x41) THEN
                  GOTO LB4        ; x21,31 = freq, x23,33 = fsc
               ELSE
                  x42 = x42 + 1
               ENDIF
            ENDIF
         ENDIF

         x21 = x31
         x23 = x33
```

```
        LB3
        LB4

x60 = x23-x33
x61 = (x40-x33)/x60
x62 = x31 - (x61 * (x31-x21)) ; x62 = interpolated spatial freq.

x66 = x65/x62   ; resolution = pixel size/spatial frequency

        DE
        [res]{***x82}

        SD /    NORM'D FREQ  CUTOFF      ANGSTROMS
        [res]{***x82}

        SD 1,x62,x40,x66
        [res]{***x82}

        UD 1,x0,x0,x67
        [res]{***x80}

        x83 = x80-1

IF (x95.EQ.1) GOTO LB8   ; Ignore resolution check?

IF (x66.GT.x67) GOTO LB5 ; Check for resolution improvement

        LB8

UD 1,x96
[ptcl_list]

IF (x80.GT.1) THEN
VM
rm [recntr_dir]/{***x83}re-ali*****
ENDIF

IF (x80.GT.1) THEN
DE
[APSH_doc]{***x83}
ENDIF

 LB1


 LB5

 DOC RAN
 [select_doc]{***x83}
 [3D_dir]/{***x83}randoc01
 (50)

 DOC SUBTRACT
 [select_doc]{***x83}
 [3D_dir]/{***x83}randoc01
 [3D_dir]/{***x83}randoc02
 (1)
```

```
 DOC REN
 [3D_dir]/{***x83}randoc02
 [3D_dir]/{***x83}randoc02

 DOC SORT
 [3D_dir]/{***x83}randoc01
 [3D_dir]/{***x83}randoc01
 (1)
 (Y)

 BP CG
 [recntr_dir]/{***x83}re-ali*****
 [select_doc]{***x83}
 x20
 [APSH_doc]{***x83}
 *
 [3D_model]{***x80}_final
 5.0E-5,0.0
 25,1
 200

 BP CG
 [recntr_dir]/{***x83}re-ali*****
 [3D_dir]/{***x83}randoc01
 x20
 [APSH_doc]{***x83}
 *
 [3D_model]{***x80}_final_odd
 5.0E-5,0.0
 25,1
 200

 BP CG
 [recntr_dir]/{***x83}re-ali*****
 [3D_dir]/{***x83}randoc02
 x20
 [APSH_doc]{***x83}
 *
 [3D_model]{***x80}_final_eve
 5.0E-5,0.0
 25,1
 200

; Create FSC curve

RF 3
[3D_model]{***x80}_final_odd
[3D_model]{***x80}_final_eve
(0.5)
(0.2,2.0)
C
x99
(3)
[FSC_doc]{***x80}_final

DOC REN
[FSC_doc]{***x80}_final
```

```
[FSC_doc]{***x80}_final

; Calculate resolution

x42 = 1 ; counter to compare with x41

UD N, x51
[FSC_doc]{***x80}_final

; Get the first line (x21 = previous, x31 = current)

UD 1,x21,x22,x23
[FSC_doc]{***x80}_final

       DO LB6 x50=1,x51

       UD x50,x31,x32,x33
       [FSC_doc]{***x80}_final

       IF (x33.LT.x40) THEN
          IF (x23.GT.x40) THEN
             IF (x42.EQ.x41) THEN
                 GOTO LB7           ; x21,31 = freq, x23,33 = fsc
              ELSE
                 x42 = x42 + 1
              ENDIF
           ENDIF
        ENDIF

        x21 = x31
        x23 = x33

        LB6
        LB7

x60 = x23-x33
x61 = (x40-x33)/x60
x62 = x31 - (x61 * (x31-x21)) ; x62 = interpolated spatial freq.

x66 = x65/x62   ; resolution = pixel size/spatial frequency

        DE
        [res]{***x80}_final

        SD /    NORM'D FREQ  CUTOFF      ANGSTROMS
        [res]{***x80}_final

        SD 1,x62,x40,x66
        [res]{***x80}_final

        EN D
```

**T3D_FSC.spi - Calculate FSC (Between 2 half models, used to determine resolution)**

```
md
tr on

;  ------------------- Parameters ------------------------------
-----
x10 = 4    ; No. classes
x11 = 1.0    ; Ring width
x12 = 0.6    ; Lower scale factor
x13 = 1.2    ; Upper scale factor
x14 = 50    ; Max. tilt angle
x15 = 3.0    ; Factor for noise comparison (in sigma)

;  ------------------- Input files ------------------------------
------

fr l
[vol_1]  ; 1st 3D model

fr l
[vol_2] ; 2nd 3D model

; -------------------------- Output files ----------------------
----

fr l
[FSC_doc] ; FSC doc file

; -------------- END BATCH HEADER ------------------------------

;DO LB1 x20=1,x10

DE [FSC_doc] ; remove any previous doc files

RF 3       ; Compute resolution curve
[vol_1]    ; First input volume
[vol_2]    ; Second input volume
x11        ; Ring width
x12,x13    ; Lower, upper scale factors
C          ; C = missing cone
x14        ; Max. tilt angle
x15        ; Factor for noise comparison
[FSC_doc]  ; Output doc file

DOC REN
[FSC_doc]
[FSC_doc]

;LB1

en d
```

**T3D_res.spi - Calculate resolution of 3D model**

```
;  ------------------- Parameters -------------------------------
```

```
x10 = 5 ; No. of classes
x40 = 0.5    ; Cutoff threshold for Fourier shell correlation curve
x41 = 1      ; Use the Nth time the curve crosses the threshold
x65 = 2.62    ; Pixel size

; -------------------- Input files --------------------------------

;FR G
;[params]../params     ; Parameter file

FR G
[combires] ; Input FSC doc file

; -------------------------- Output files -------------------------
FR G
[res]  ; Doc file with resolution data

; -------------- END BATCH HEADER --------------------------------

;DO LB3 x20=1,x10 ; loop thro classes

DE [res] ; remove any previous res files

x42 = 1 ; counter to compare with x41

UD N, x51
[combires]

UD 1,x21,x22,x23     ; Get the first line (x21 = previous, x31 =
current)
[combires]

DO LB1 x50=1,x51

UD x50,x31,x32,x33
[combires]

IF (x33.LT.x40) THEN
   IF (x23.GT.x40) THEN
      IF (x42.EQ.x41) THEN
         GOTO LB2           ; x21,31 = freq, x23,33 = fsc
      ELSE
         x42 = x42 + 1
      ENDIF
   ENDIF
ENDIF

x21 = x31
x23 = x33
LB1
LB2

x60 = x23-x33
x61 = (x40-x33)/x60
x62 = x31 - (x61 * (x31-x21)) ; x62 = interpolated spatial freq.

;UD 5,x65    ; pixels size
;[params]

x66 = x65/x62   ; resolution = pixel size/spatial frequency
```

```
DE
[res]

SD /    NORM'D FREQ  CUTOFF      ANGSTROMS
[res]

SD 1,x62,x40,x66
[res]

;LB3

EN D
;
```

**T3D_3D-filt.spi - Filter 3D model**

```
; --------------- Parameters ---------------------------------------

;x12 = 0.01 ; PASS BAND
;x10 = 5 ; No. of classes
x14 = 0.05    ; Lower freq. limit (default is 0.05)
x15 = 0.05    ; Upper freq. limit (default is 0.05)

; --------------- Input files --------------------------------------
-

FR G
[res] ; Doc file with resolution data

FR G
[vol] ; model to filter

; --------------- Output files -------------------------------------

FR G
[fvol] ; Filtered model

; -------------- END BATCH HEADER ---------------------------------

;DO LB1 x20=1,x10 ; loop thro classes

UD 1,x11          ; Get cutoff frequency value from resolution doc
file
[res]

x12 = x11 - x14  ;  Lower freq. limit, default is 0.05
x13 = x11 + x15  ;  Upper freq. limit, default is 0.05

FQ                ; Filter quick
[vol]             ; Input volume
[fvol]            ; Output volume
(7)                 ; Filter type; 7 = Butterworth low-pass
x12,x13           ; Upper, lower frequency limits

;LB1

EN D
```